

Psychological Review

THEODORE M. NEWCOMB, Editor

UNIVERSITY OF MICHIGAN

Lorraine Bouthilet, Managing Editor

CONTENTS

David Katz 1884-1953.....	ROBERT B. MACLEOD	1
The Physiology of Motivation.....	ELIOT STELLAR	5
The S-R Reinforcement Theory of Extinction.....	HENRY GLEITMAN, JACK NACHMIAS, & ULRIC NEISSER	23
Punishment: I. The Avoidance Hypothesis.....	JAMES A. DINSMOOR	34
The Measurement of Values.....	L. L. THURSTONE	47
A Neural Model for Sign-Gestalt Theory.....	JAMES OLDS	59
The Place of Physiological Constructs in a Genetic Explanatory System.....	GUDMUND SMITH	73
A Note on Stimulus Intensity Dynamism (<i>V</i>).....	FRANK A. LOGAN	77

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

CONSULTING EDITORS

SOLOMON ASCH
ROBERT BLAKE
STUART W. COOK
CLYDE COOMBS
LEON FESTINGER
W. R. GARNER
JAMES J. GIBSON
D. O. HEBB
HARRY HELSON
E. R. HILGARD
CARL I. HOVLAND
E. LOWELL KELLY
DAVID KRECH
ROBERT W. LEEPER

ROBERT B. MACLEOD
DAVID C. MCCLELLAND
G. A. MILLER
GARDNER MURPHY
OSCAR OESER
CARROLL C. PRATT
DAVID SHAKOW
RICHARD SOLOMON
ELIOT STELLAR
S. S. STEVENS
ERIC TRIST
EDWARD WALKER
ROBERT WHITE

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

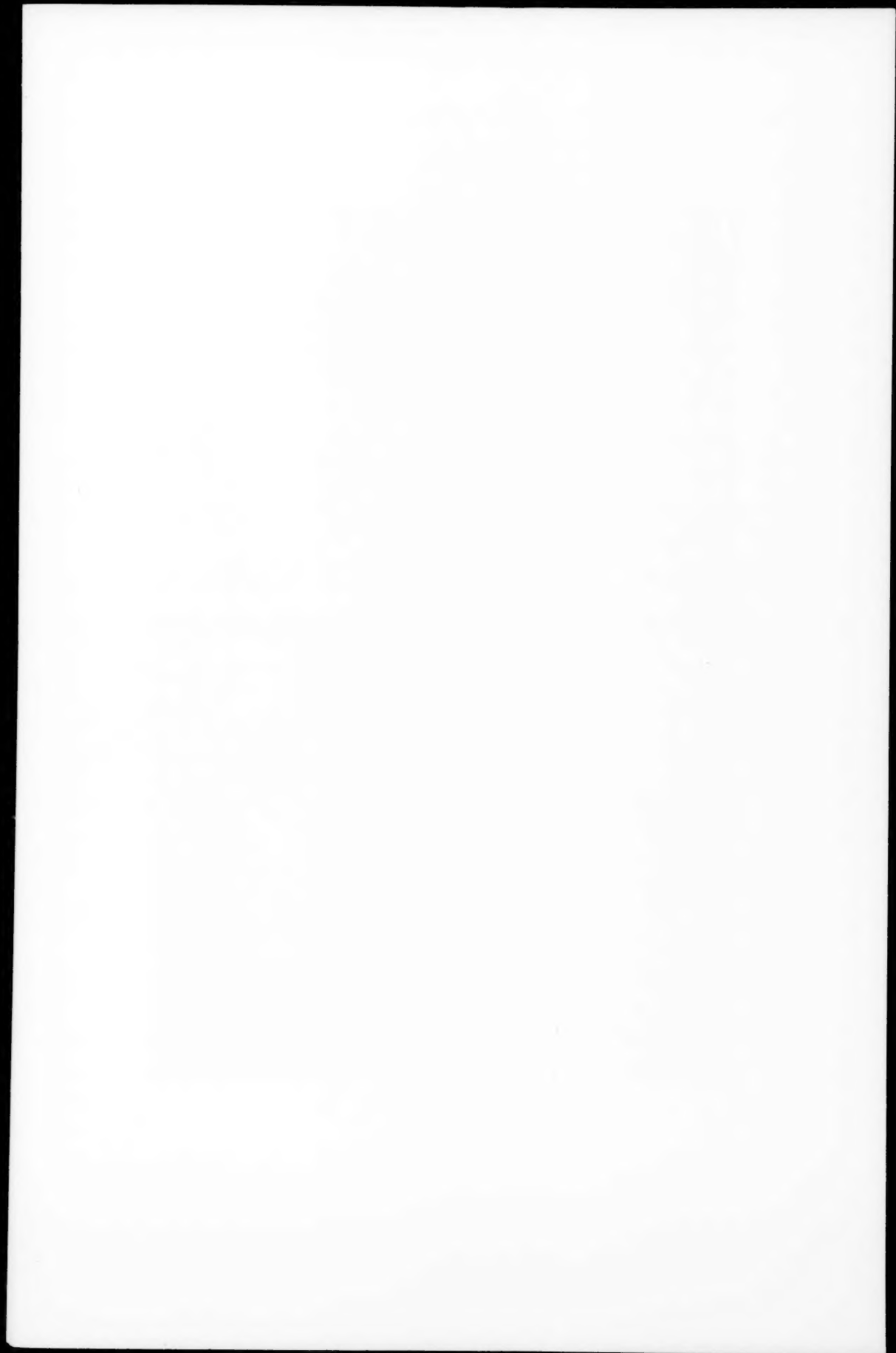
\$6.50 volume

\$1.25 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Copyright 1954 by the American Psychological Association, Inc.





DAVID KATZ

THE PSYCHOLOGICAL REVIEW

DAVID KATZ

1884-1953

David Katz, Professor Emeritus at the University of Stockholm, died of a sudden heart attack on February 2, 1953. By those who attended the International Congress of Psychology in Stockholm in July, 1951, he will be remembered as the indefatigable organizer and genial host of the congress. In the history of psychology his name will be associated with significant contributions to almost every field of psychology, pure and applied; and he will be cited as one of this century's outstanding exponents of psychological phenomenology. In the memories of those who knew him and loved him he will live as a gentle, humble man, persistently curious about everything that had to do with human nature, brilliant in his intuitions, tireless in his research, unfailingly generous and courteous in controversy.

Katz was born in Kassel, Germany, on October 1, 1884. His early education was in Kassel, his university education in Berlin, Munich, and Göttingen, where he received his doctoral degree in 1906. In Göttingen he was one of G. E. Müller's most brilliant pupils. Later he became Müller's assistant, and, in 1911, Privat Dozent. During World War I he was called to army service for four years, returning afterwards to his post in Göttingen. It was during his Göttingen period that he completed his now classic researches on the experimental phenomenology of color, and began his less well-known but equally significant work on touch.

In 1919 he accepted the chair of psychology and education at the University of Rostock, where he developed what eventually became one of the most productive psychological laboratories in Europe. In 1933 the National Socialist party came into power, and Katz, as a non-Aryan, was deprived of his position. Fortunately his British friends were willing to provide hospitality and, for the next four years, first in Manchester and later in London, he was able to pursue his scientific work. In 1937 he accepted the chair of education (including psychology) at the University of Stockholm, where he remained until his retirement in 1952.

Katz paid two visits to the United States, in 1929 as Visiting Professor at the University of Maine and in 1950 as Hitchcock Lecturer at the University of California.

During his period as G. E. Müller's assistant, Katz was fond of relating, an attractive young Russian girl was admitted as a student. In reply to Katz's query, Müller characterized her as "eine Madonna mit einer Bombe." Rosa Heine did not blow up the Institute, thereby failing to conform to Müller's stereotype of the Russian, but she speedily conquered Müller's assistant. Katz and Rosa Heine were married in 1919. Numerous joint publications attest to their productivity as a scientific team. Their two sons, now launched on their own professional careers, were made prematurely famous by their

parents' book, *Gespräche mit Kindern* (1927).

To review Katz's contributions to psychology would be a major undertaking. As a scientist he had "green fingers." He had but to touch a problem, and it readily blossomed and bore fruit. His list of publications includes more than 100 titles, of which at least 20 are substantial books and monographs. Among these one finds contributions to animal, child, educational, abnormal, and social psychology, to the experimental psychology of perception, motivation, learning, and thinking, to systematic theory, and to laboratory instrumentation. It may be that he scattered his energies too widely; certainly, not all his researches are of equal merit. It was his genius, however, to find in the commonplace observations of daily life problems which, when viewed in a larger context, became significant, and to make psychological capital out of every new experience with which good or bad fortune provided him. Thus, his wartime assignment to a military hospital led to a pioneer study of the psychological problems of amputees, and later to the invention of a device for the training of students in the technique of percussion; the feeding problems of his children contributed to his interest in constitutional typology and in the theory of hunger and appetite; his own difficulty with the English and Swedish languages challenged him as a psychologist to do some experiments on problems of language and thinking.

It was also his genius to find simple and inexpensive ways of attacking major problems. Katz belonged perforce to the cardboard and thumbtack school; but he never allowed a meager budget to hamper his activity. In Rostock he was faced with the task of developing a research institute on an annual budget of approximately \$125. Some of his problems required the use of animals.

He could not afford a regular animal laboratory; so he bought some chicks. Out of his chicken yard came the well-known *Hackgesetz*, the studies of chickens reared in isolation, the studies of "counting" behavior in chickens, and the experiments that led to the "avidity" theory of appetite. While in England, lacking an adequate laboratory, he pursued his tactual researches by undertaking some assignments for the flour millers, who were concerned about the elasticity of their dough. When he arrived in Sweden, he was assigned a small apartment as a laboratory. The kitchen promptly became a workshop, the bathroom became a photographic darkroom, a fifteen-year-old boy served as technician, and with cardboard, thumbtacks, bathroom scales, and sticks of wood, the laboratory began to produce research. When one thinks of David Katz, one wonders sometimes whether handsome budgets are a hindrance or an aid to productivity.

The frustrated graduate student in search of a doctoral problem has but to thumb through a few of Katz's publications to find a wealth of inviting questions and challenging hypotheses that will draw him straight to the laboratory. The human hand as a unitary sense organ analogous to the eye, the composite photograph as a device for the study of group characteristics, the sensory basis of the phenomenon of elasticity, the phantom limb of the amputee, the ability of certain deaf people to appreciate music, and a host of other apparent byways of psychological investigation were opened up by Katz and redirected towards the central problem. It was characteristic of his restless curiosity, however, that he was frequently content to blaze the trail, bequeathing to another generation the task of exploiting it.

The unity within Katz's apparent diversity of interest is to be found in his

consistent application of the phenomenological method. He was interested in the prediction and control of behavior, in the social and biological determinants of behavior, in the tricky problems of instrumentation, in the broader problems of psychological theory, but behind it all was a persistent, a passionate curiosity about the world of phenomena. For Katz the most fascinating thing to wonder about was a human experience. It might be a simple color or sound, or the strange beauty of an El Greco picture, or the peculiar sensations that accompany the crunching of a nut between the teeth, or the ineffable satisfyingness of a cool draught of beer on a warm day. All experience was something to appreciate and to wonder about. For him the first task of the psychologist—not really a task, but a pleasure—was to observe and describe without bias both the salient characteristics and the subtle nuances of ordinary human experience. Phenomenology for him was essentially an attitude of “disciplined naiveté.” From descriptive analysis one proceeds to experiment and to theory, but no psychological theory, he argued, could be complete if it excluded any of the essential variables of human experience.

Katz's psychological phenomenology is best exemplified in his studies of color and touch, *Die Erscheinungsweisen der Farben* (1911¹) and *Der Aufbau der Tastwelt* (1925). Influenced by the physiologist Hering and the philosopher Husserl he insisted that the psychologist should begin by deliberately “bracketing” his physical, physiological, and philosophical biases and attempt to observe phenomena as they are actually presented. The phenomenal world thus viewed contains properties and relationships that escape the notice of the phys-

ically or physiologically oriented observer. The classical psychologist was content to order colors in terms of hue, brightness, and saturation; Katz saw them also varying in mode of appearance, pronouncedness, insistence, transparency, inherence, and stability. Classical psychology was busily mapping the patterns of pressure, pain, warm, and cold spots on the skin, and searching for receptors; Katz went further, and explored the active process of “touching” (*tasten*), discovering here, too, modes of appearance, properties of organization, and unsuspected kinds of sensitivity. It is unfortunate that, while his visual studies have been widely appreciated, his richly suggestive book on the world of touch has received relatively little notice.

During recent years the word *phenomenological* has tended to expand its meaning. It is coming to suggest an easy-going, intuitive, sympathetic “seeing the world as the other fellow sees it,” an approach that permits one to take things at their face value and to avoid the rigors of experimentation and theory construction. This is definitely not the kind of psychological phenomenology that Katz advocated. True, he was interested in the “fuzzy” aspects of experience; but for him the “fuzziness” of a phenomenon was no excuse for careless observation or undisciplined thinking. Good phenomenology, he held, requires at least as much training and discipline as does good Titchenerian introspection. Nor does phenomenology lead away from experimentation and theory; it is an essential first step in the direction of more imaginative experimentation and sounder theory.

Katz adhered to no “school” of psychology, nor—which is strange in a German of his generation—did he ever attempt to found a school. In his sympathies he stood closest to the Gestalt theorists; indeed, his pioneer work on

¹ Later revised as *Der Aufbau der Farbwelt* (1930); abridged and translated into English as *The World of Colour* (1935).

phenomenal constancy must be regarded as basic to the Gestalt theory of perception, and his more recent experiments on thinking belong in the Gestalt tradition. His interests were too varied, however, to fit neatly within any formal system, and we find him in his *Gestaltpsychologie* (1944) expressing impatience with the narrowness of the Gestalt approach. Like Stern he believed that every part process must be understood ultimately in terms of the total person, but he lacked Stern's compulsion to turn his personalism into a philosophy. With Jaensch he shared an interest in the possibilities of typology, but for him typology was a problem for research rather than a revelation. He found merit in the developmental approach, both ontogenetic and phylogenetic, but he rejected the extremes of both nativism and empiricism. He was willing to accept

physiological evidence and to do physiological experiments when he felt that such would help to clarify a psychological problem, but he refused to accord to physiological constructs any unique explanatory value.

It is perhaps best to think of Katz as essentially a pioneer, catholic rather than eclectic, ready to adapt to his purposes any tool, material or conceptual, that looks useful, but never forgetting the purpose for which he has selected it. For Katz there was a single purpose that persisted throughout his scientific life. It was, to put it in old-fashioned language, to understand the phenomena of the human mind. Those who see as a challenge to science all the phenomena of human mentality will find in Katz a kindred spirit.

ROBERT B. MACLEOD

Cornell University

THE PHYSIOLOGY OF MOTIVATION

ELIOT STELLAR

The Johns Hopkins University

In the last twenty years motivation has become a central concept in psychology. Indeed, it is fair to say that today it is one of the basic ingredients of most modern theories of learning, personality, and social behavior. There is one stumbling-block in this noteworthy development, however, for the particular conception of motivation which most psychologists employ is based upon the outmoded model implied by Cannon in his classical statement of the local theories of hunger and thirst (23). Cannon's theories were good in their day, but the new facts available on the physiological basis of motivation demand that we abandon the older conceptualizations and follow new theories, not only in the study of motivation itself, but also in the application of motivational concepts to other areas of psychology.

This argument for a new theory of motivation has been made before by Lashley (42) and Morgan (47). But it is more impelling than ever today because so much of the recent evidence is beginning to fit into the general theoretical framework which these men suggested. Both Lashley and Morgan pointed out that the local factors proposed by Cannon (e.g., stomach contractions or dryness of the throat) are not necessary conditions for the arousal of motivated behavior. Instead, they offered the more inclusive view that a number of sensory, chemical, and neural factors cooperate in a complicated physiological mechanism that regulates motivation. The crux of their theory was described most recently by Morgan as a *central motive state (c.m.s.)* built up in the organism by the combined influences of the sensory, humoral, and neu-

ral factors. Presumably, the amount of motivated behavior is determined by the level of the *c.m.s.*

Beach (8, 11), in his extensive work on the specific case of sexual motivation, has amply supported the views of Lashley and Morgan. But the important question still remains: Do other kinds of motivated behavior fit the same general theory? As you will see shortly, a review of the literature makes it clear that they do. As a matter of fact, there is enough evidence today to confirm and extend the views of Lashley, Morgan, and Beach and to propose, in some detail, a more complete physiological theory of motivation.

There are a number of ways to present a theoretical physiological mechanism like the one offered here. Perhaps the best approach is to start with an overview and summarize, in a schematic way, the major factors at work in the mechanism. Then we can fill in the details by reviewing the literature relevant to the operation of each factor. Some advantage is lost by not taking up the literature according to behavioral topics, that is, different kinds of motivation. But the procedure adopted here lets us focus attention directly on the theory itself and permits us to make some very useful comparisons among the various kinds of motivation. Once the theoretical mechanism and the evidence bearing on it are presented, the final step will be to evaluate the theory and show what experiments must be done to check it and extend it.

THEORETICAL SCHEME

A schematic diagram of the physiological mechanism believed to be in con-

trol of motivated behavior is shown in Fig. 1. The basic assumption in this scheme is that *the amount of motivated behavior is a direct function of the amount of activity in certain excitatory centers of the hypothalamus*. The activity of these excitatory centers, in turn, is determined by a large number of factors which can be grouped in four general classes: (a) *inhibitory hypothalamic centers* which serve only to depress the activity of the excitatory centers, (b) *sensory stimuli* which control hypothalamic activity through the afferent impulses they can set up, (c) *the internal environment* which can influence the hypothalamus through its rich vascular supply and the cerebrospinal fluid, and (d) *cortical and thalamic centers* which can exert excitatory and inhibitory influences on the hypothalamus.

As can be seen, the present theory holds that the hypothalamus is the seat of Morgan's *c.m.s.* and is the "central nervous mechanism" Lashley claimed was responsible for "drive." Identifying the hypothalamus as the main integrating mechanism in motivation makes the experimental problem we face more

specific and more concrete than ever before. But it also makes it more complicated, for the physiological control of the hypothalamus is exceedingly complex. The influence of the internal environment on the hypothalamus is changing continuously according to natural physiological cycles, and of course it may often be changed directly by the chemical and physical consequences of consummatory behavior (see Fig. 1). Sensory stimuli may also have varied effects on the hypothalamic mechanism, depending upon their particular pattern, previous stimulation, previous learning, sensory feedback from the consummatory behavior itself, and the influence the internal environment has already exerted on the hypothalamus. Similarly, the influence of the cortex and thalamus will add to the hypothalamic activity already produced by sensory stimuli and the internal environment. Presumably, these cortical and thalamic influences may result directly or indirectly from sensory stimulation, but they may also be controlled partly by the "upward drive" of the hypothalamus itself (43). Then, to complicate the picture even more, there are the inhibitory centers of the hypothalamus which are also controlled by the various internal changes, sensory stimuli, and cortical and thalamic influences. These centers, presumably, depress the activity of the excitatory centers and, therefore, attenuate their output.

Fortunately, this mechanism is not as formidable against experimental attack as it might appear. The basic experimental approach is to isolate the controlling factors in any type of motivation and determine their relative contributions to hypothalamic activity. As you will see, a number of experimental techniques like sensory deprivation, hormone and drug administration, cortical ablation, and the production of subcortical lesions may be used fruitfully

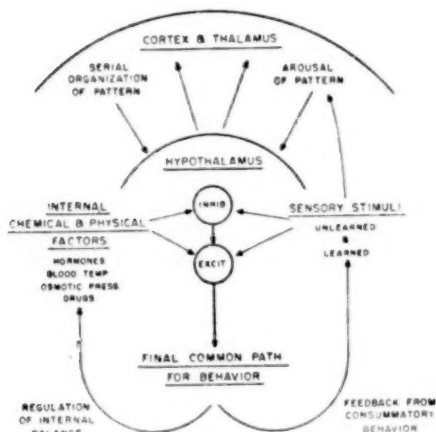


FIG. 1. Scheme of the physiological factors contributing to the control of motivated behavior. (See text.)

to isolate these factors. But that is only half the problem. Obviously, the factors controlling hypothalamic activity and motivation do not operate in isolation. In fact, it is quite clear that their influences interact. Therefore, it becomes an equally important problem to determine the relative contribution of each factor while the others are operating over a wide range of variation.

EXPERIMENTAL EVIDENCE

Before going into the literature bearing on the operation of each of these factors in control of motivated behavior, it will help to raise a few questions that ought to be kept in mind while considering the experimental evidence. Are there different hypothalamic centers controlling each kind of motivation? Does the hypothalamus exert its influence through direct control of the final effector pathways or does it simply have a "priming" effect on effector paths controlled by other parts of the nervous system? Do all these factors operate in the control of each type of motivation or are there cases where sensory stimuli, for example, may not be important or where changes in the internal environment do not contribute? Can the same mechanism describe the control of motivation measured by simple consummatory behavior, preference, and learning? Are the same mechanisms involved in the control of simple, biological motives and complex, learned motives?

Hypothalamic centers. Review of the literature on the role of the hypothalamus in motivation brings out three general conclusions. (a) Damage to restricted regions of the hypothalamus leads to striking changes in certain kinds of motivated behavior. (b) Different parts of the hypothalamus are critical in different kinds of motivation. (c) There are both excitatory and inhibitory

centers controlling motivation in the hypothalamus; that is, damage to the hypothalamus can sometimes lead to an increase in motivation and sometimes a marked decrease.

The evidence bearing on these three points can be summarized briefly. Many experiments have shown that restricted bilateral lesions of the hypothalamus will make tremendous changes in basic biological motivations like hunger (16, 22), sleep (49, 50, 53), and sex (6, 18, 20). Less complete evidence strongly suggests that the same kinds of hypothalamic integration is also true in the cases of thirst (61), activity (35), and emotions (5, 62). We have only suggestive evidence in the case of specific hungers (59).

It is clear that there is some kind of localization of function within the hypothalamus although it is not always possible to specify precisely the anatomical nuclei subserving these functions. The centers for hunger are in the region of the ventromedial nucleus which lies in the middle third of the ventral hypothalamus, in the tuberal region (16). (See Fig. 2.) Sleep is controlled by centers in the extreme posterior (mammillary bodies) and extreme anterior parts of the hypothalamus (49, 50). The critical region for sexual behavior is in the anterior hypothalamus, between the optic chiasm and the stalk of the pituitary gland (18, 20). The center for activity is not clearly established, but seems to be adjacent with or overlapping the centers for hunger (35). Finally, the centers for emotion are also in the vicinity of the ventromedial nucleus, perhaps somewhat posterior to the hunger centers and overlapping the posterior sleep center (50, 62).

In at least two cases it is clear that there must be both excitatory and inhibitory centers controlling motivated behavior. In the case of hunger, bilateral lesions in the ventromedial nucleus

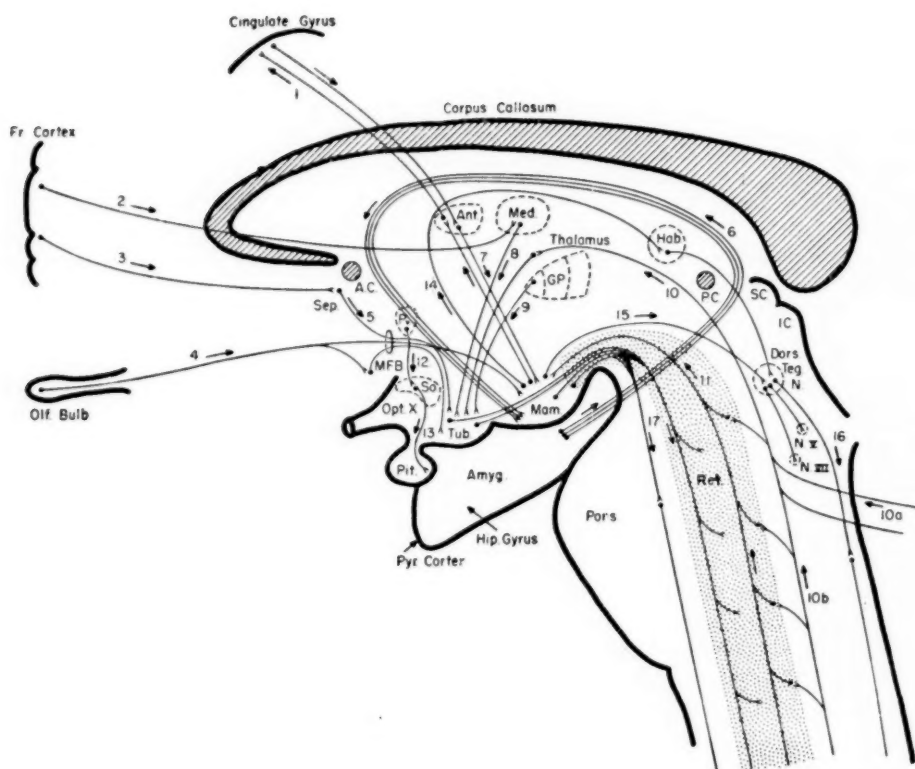


FIG. 2. Schematic drawing of the hypothalamus and its major neural connections. Adapted from W. R. Ingram's diagram in Gellhorn (30) and D. B. Lindsley's Figure 9 (43).

Abbreviations and Description of Pathways

A.C.	Anterior commissure
Amyg.	Amygdala
Ant.	Anterior thalamic nuclei
Cingulate Gyrus	Cortex of cingulate gyrus
Dors. Teg. N.	Dorsal tegmental nucleus
Fr. Cortex	Cortex of frontal lobe
GP	Globus pallidus
Hab.	Habenular nucleus of thalamus
Hip. Gyrus	Hippocampal gyrus
IC	Inferior colliculus
Mam.	Mammillary nuclei
Med.	Dorsal medial thalamic nucleus
MFB	Medial forebrain bundle
N.V	Motor nucleus, Vth nerve
N.VII	Motor nucleus, VIIth nerve
Olf. Bulb	Olfactory bulb
Opt. X	Optic chiasm
P.C.	Posterior commissure
Pit.	Pituitary gland
Pv.	Paraventricular nucleus
Pyr. Cortex	Pyriform cortex
Ret.	Reticular formation
SC	Superior colliculus
Sep.	Septal nuclei
So.	Supraoptic nucleus
Tub.	Tuber cinereum

Afferents to Hypothalamus

1. Corticothalamic fibers
2. Frontothalamic fibers
3. Frontoseptal fibers
4. Olfacto-hypothalamic tract
5. Septo-hypothalamic fibers
6. Fornix
7. Mammillothalamic tract
8. Thalamo-hypothalamic fibers
9. Pallido-hypothalamic fibers
10. Sensory systems ascending to thalamus
 - 10 a. cranial afferents
 - 10 b. somatic and visceral afferents
11. Sensory collaterals to hypothalamus
12. Paraventriculo-supraoptic fibers

Efferents from Hypothalamus

13. Supraoptic hypophyseal tract
14. Mammillohabenular tract
15. Mammillotegmental tract
16. Dorsal longitudinal fasciculus
17. Descending efferents relaying in brain stem and medulla

near the midline produce a tremendous amount of overeating (3, 16). Such a center is presumably an inhibitory one since removing it leads directly to an

increase in eating behavior. On the other hand, lesions $1\frac{1}{2}$ to 2 millimeters off the midline at the level of the ventromedial nucleus completely eliminate

hunger behavior (3, 4). After such lesions animals never eat again, so we can call such centers excitatory centers. Supporting this interpretation is the fact, recently reported, that stimulating these lateral centers in the waking cat through implanted electrodes results in vast overeating (27). The same sort of mechanism turns up in the case of sleep. In the posterior hypothalamus, in the region of the mammillary bodies, there are excitatory centers or "waking" centers which operate to keep the organism awake (49, 50). When they are removed, the animal becomes somnolent and cannot stay awake. In the anterior hypothalamus, around the pre-optic nucleus, there is an inhibitory center (49). When that is removed, the animal is constantly wakeful.

So far, only an excitatory center has been found in the case of sexual behavior. Bilateral lesions anterior to the pituitary stalk eliminate all mating behavior (18, 20), but no lesion of the hypothalamus has ever been reported that resulted in an exaggeration of sexual motivation. What little we know about the center for activity near the ventromedial nucleus suggests that it is also an excitatory center since lesions there produce only inactivity and not hyperactivity (35). In the case of emotions, the picture is not yet clear. Lesions near the ventromedial nucleus make cats highly emotional (62), and therefore this center must be inhibitory. But the lateral regions of the posterior hypothalamus seem to be excitatory, for lesions there make animals placid (50). Furthermore, direct stimulation of these posterior regions produces many of the signs of rage reactions (52).

There is some evidence that sheds light on how the excitatory and inhibitory hypothalamic centers may cooperate in the regulation of motivation. In the clear-cut cases of sleep and hunger it appears that the inhibitory centers

operate mainly through their effects on the excitatory centers. At least we know that when both centers are removed simultaneously the effect is indistinguishable from what happens when only the excitatory centers are removed (3, 49). So it is convenient for present theoretical purposes to think of the inhibitory center as one of the factors which influences the level of activity of the excitatory center. In fact, to speculate one step further, it is worth suggesting that the inhibitory centers may constitute the primary neural mechanism regulating the satiation of motivation.

Sensory stimuli. What effects do sensory stimuli have upon the hypothalamus and how important are such stimuli in the control of motivation? Some answer to the first part of this question is given by the schematic outline of hypothalamic connections shown in Fig. 2. Clearly the hypothalamus has a rich supply of afferents coming directly or indirectly from all the various sense organs. In fact the diagram is really an understatement of hypothalamic connections because it is an oversimplified and conservative representation. Physiological evidence shows, for example, that there must be connections from the taste receptors via the solitary nucleus of the medulla (36). Also there is evidence of rich connections from the visual system via the lateral geniculate of the thalamus (36). There is no doubt about the fact that the hypothalamus is under very extensive sensory control.

As to the sensory control of motivation, there is excellent reason to believe that the stimuli which can set up impulses in these pathways to the hypothalamus are of particular importance. Perhaps the best example comes from the study of sexual behavior (11). The consensus of a group of studies on different mammals is as follows. Sexual behavior is not dependent upon any single sensory system. Extirpation of any

one peripheral sense organ has no appreciable influence on the arousal and execution of sexual behavior. If two sensory avenues are destroyed, however, sexual behavior may be eliminated, especially in the case of the naive animal. With experienced animals, interestingly enough, it may take destruction of three sensory systems. But in neither case does it matter what combination of sensory systems is eliminated. We can conclude, therefore, that it is the sum total of relevant sensory impulses arriving at the central nervous system (hypothalamus) that is important in setting off sexual behavior.

Kleitman's analysis of sleep and wakefulness shows that the same kind of sensory control operates in this case (38). Wakefulness seems to be dependent upon the sum total of sensory impulses arriving at the waking center in the posterior hypothalamus, regardless of the particular sensory systems involved. Direct support of this kind of view is offered by Bremer's (14) physiological data which showed that maintenance of the waking rhythm of the brain is less a matter of any particular sensory input and more a matter of the amount of sensory input.

What we know about hunger and thirst suggests that the amount of motivated behavior in these cases should be a joint function of sensory impulses arising from gastric contractions or dryness of the throat and taste, tactile, and temperature receptors in the mouth. Unfortunately we have no sensory deprivation experiments that are a good test of this point. But all the evidence on the acceptability of foods and fluids of different temperatures, consistencies, and flavoring suggests the joint operation of many stimuli in the control of these types of motivation.

So far, we have mentioned only stimuli which arouse motivation. What stimulus changes could reduce motiva-

tion and perhaps lead to satiation? There are three general possibilities: (a) a reduction in excitatory stimuli, (b) interfering or distracting stimuli that elicit competing behavior, and (c) "inhibitory" stimuli. It is easy to find examples of the first two types of stimulus changes and to guess their mechanisms of operation in terms of the present theory. In the case of "inhibitory" stimuli, however, all we have is suggestive evidence. For example, the fact that dogs with esophageal fistulas eat (37) and drink (1, 13) amounts proportional to the severity of deprivation suggests that the stimuli which feed back from consummatory behavior might have a net inhibitory effect on motivation (see Fig. 1). Furthermore, some of the experiments on artificially loading the stomach suggest that a full gut may result in stimuli which inhibit further eating (37) or drinking (2, 13) over and above the possibility that there might be no room left in the stomach or that gastric contractions are reduced.

In summary, we can state the following working hypotheses about the sensory factors which operate in the control of motivation. (a) No one sensory avenue is indispensable in the arousal of motivated behavior. Instead, sensory stimuli have an additive effect on the excitability of the hypothalamus so that it is the sum total of relevant impulses arriving at the excitatory centers of the hypothalamus that determine the amount of motivated behavior. (b) Judging from the resistance of experienced animals to the effects of sensory deprivation in the case of sexual motivation, it seems clear that excitatory influences in the hypothalamus may be exerted by learned as well as unlearned stimuli. (c) There are afferent impulses to the hypothalamus which have a net inhibitory effect on the excitatory centers and thus serve to reduce motivation

or produce satiation. The best guess at present is that these "inhibitory" stimuli operate by exerting an excitatory influence on the inhibitory centers of the hypothalamus. Presumably, impulses to inhibitory centers have the same kind of additive properties as impulses to the excitatory centers.

Internal environment. That the internal environment plays an important role in certain kinds of motivated behavior is a well-established fact. Two basic questions must be asked, however, before we can understand much about how the internal environment does its work. What kinds of changes that can occur in the internal environment are the important ones in motivation? How do changes in the internal environment influence the nervous system and, therefore, motivated behavior?

In terms of the present theory, we would expect the internal environment to operate in motivation by changing the excitability of hypothalamic centers. This is a reasonable expectation, for the hypothalamus is the most richly vascularized region of the central nervous system (24). Not only that, but the hypothalamus is also in direct contact with the cerebrospinal fluid in the third ventricle.

The case of sexual behavior again makes an excellent example. Experiments on the spayed, female cat (6, 17) and spayed, female guinea pig (28) have shown that hypothalamic regions must be intact and functioning if injected sex hormones are to arouse estrous behavior. If a section is made through the spinal cord only rudimentary fragments of sexual behavior can be elicited by appropriate stimulation, and injected sex hormones make no contribution to the response. Essentially the same thing is true if the section is made high in the hind brain but excludes the hypothalamus. When the decerebration is just above the hypo-

thalamus, full estrous reactions can be aroused by appropriate stimulation, but only if sex hormones have been administered. It is clear, then, that not only is the hypothalamus the main integrating center for sexual reactions, but it is also most likely the main site of action of the sex hormones. This point is further supported by studies of female guinea pigs with pinpoint lesions of the anterior hypothalamus. These animals fail to show sexual behavior even under the influence of massive doses of sex hormones (19).

A very similar mechanism seems to be involved in the case of motivated behavior dependent upon the organism's defenses against temperature extremes (activity, nesting, hoarding, selection of high-calorie diets). We know, for example, that reactions regulating body temperature in the face of heat and cold are integrated in two separate centers in the hypothalamus (15, 51). Lesions in the anterior hypothalamus destroy the ability to lose heat and, therefore, to survive in high temperatures. Posterior hypothalamic lesions, conversely, result in a loss of heat production mechanisms so that the animal succumbs to cold. Furthermore, artificially raising the temperature of the anterior hypothalamus will quickly induce heat loss, suggesting that normally the temperature of the blood may be important in activating the hypothalamic mechanisms (15, 44). Unfortunately our information stops here. There are no direct physiological studies on the role of these temperature-regulating mechanisms in the control of motivated behavior like activity, hoarding, nesting, or food selection. But it seems clear that the temperature of the blood may be one of the kinds of changes in the internal environment that can affect the hypothalamus, and it may be important in motivated behavior.

Ample evidence demonstrates that there are important changes in the internal environment involved in other kinds of motivated behavior. In hunger it has been shown that chemicals like insulin (32, 33, 48) and d-amphetamine (57) influence the rate of eating. It is clear that these chemicals do not operate primarily through their effects on gastric contractions, but it is only by a process of elimination that we can guess that their sites of action are in the hypothalamus. Supporting this possibility is the evidence that there are chemoreceptors in the hypothalamus which are sensitive to variations in blood sugar and important in the regulation of hunger (45). In the case of specific hungers, much evidence shows that food preference and diet selection depend upon changes in the internal environment produced by such things as pregnancy, dietary deficiencies, or disturbances of endocrine glands (54). Furthermore there are some preliminary experimental data, in the case of salt and sugar appetites, to suggest that there are separate regulatory centers in the hypothalamus which are responsive to changes in salt and sugar balance (59). Finally, in the case of thirst we know that a change in osmotic pressure, resulting from cellular dehydration, is the important internal change leading to drinking behavior (31). We know further that in the hypothalamus there are nerve cells, called "osmoreceptors," which are extremely sensitive to minute changes in osmotic pressure (61). But the direct experiment has not been done to check whether or not it is these nerve cells which are mainly responsible for the control of thirst.¹

¹ In a recent publication, Anderson of Stockholm has shown that injection of small quantities of hypertonic NaCl directly into restricted regions along the midline of the hypothalamus produces immediate and extensive drinking in water-satiated goats. (Anderson,

Obviously the experimental evidence on hunger, specific hunger, and thirst is incomplete. But enough of it fits into the scheme of the theoretical mechanism proposed here to suggest the real possibility that the internal changes important in these cases operate largely through their effects on the hypothalamus.

One question still remains. What role does the internal environment play in the mechanism of satiation? About all we have to go on at present is the very striking fact from the case of specific hungers that vastly different amounts of consummatory behavior are needed to bring about satiation for different food substances. In vitamin deficiencies only a few milligrams of substance need be consumed to produce satiation, whereas in caloric deficiencies many grams of carbohydrate, fat, or protein must be ingested. Presumably, it is not the sensory feedback from consummatory behavior that is important in these cases, but rather some inhibitory effects produced by what is consumed (Fig. 1). Within the present theoretical framework, such inhibitory effects could be produced either by depression of excitatory centers of the hypothalamus or by arousal of activity in inhibitory centers. The problem is an important one and it is wide open for study.

It is clear from the foregoing that many types of motivated behavior are dependent upon changes in the internal environment. Several points are worth emphasizing. (a) A variety of kinds of changes in the internal environment can play a role in the regulation of motivation: variation in the concentration of certain chemicals, especially hormones, changes in osmotic pressure, and

B. The effect of injections of hypertonic NaCl-solutions into different parts of the hypothalamus of goats. *Acta Physiol. Scand.*, 1953, 28, 188-201.)

changes in blood temperature. (b) The best hypothesis at present is that these internal changes operate by contributing to the activity of excitatory hypothalamic centers controlling motivation. (c) An equally important but less well-supported hypothesis is that internal changes, normally produced by consummatory behavior, operate in the production of satiation by depressing excitatory centers or arousing inhibitory centers of the hypothalamus.

Cortical and thalamic centers. Despite the heavy emphasis laid upon the hypothalamus in this discussion, it is obvious that it is not the only neural center operating in the control of motivated behavior. In the first place, some of the sensory, motor, and associative functions of the cortex and thalamus are directly important in motivation quite apart from any influence they have on the hypothalamus. Secondly, even though the hypothalamus may be the main integrating center in motivation, it does not operate in isolation. There is much evidence that the hypothalamus is under the direct control of a number of different cortical and thalamic centers (Fig. 2).

The case of emotions offers the best example of how the cortex may operate in motivation. According to the early work of Bard and his co-workers on the production of "sham rage" by decortication, it looked as though the entire cortex might normally play an inhibitory role in emotions (5). More recent work, however, shows that cortical control of emotion is more complicated than this. Bard and Mountcastle (7), for example, have found that removal of certain parts of the old cortex (particularly amygdala and transitional cortex of the midline) produced a tremendous increase in rage reactions in cats. On the other hand, removing only new cortex resulted in extremely placid cats. Results of work with monkeys (40) and

some very recent experiments with cats disagree somewhat with these findings in showing that similar old cortex removals lead to placidity rather than ferocity. The disagreement is yet to be resolved, but at least it is clear that different parts of the cortex may play different roles in the control of emotion, certain parts being inhibitory and others excitatory.

In the case of sleep, it appears so far that the cortex and thalamus play excitatory roles, perhaps having the effect of maintaining the activity of the waking center in the posterior hypothalamus. Decortication in dogs, for example, results in an inability to postpone sleep and remain awake for very long, or, as Kleitman puts it, a return to polyphasic sleep and waking rhythms (38, 39). Studies of humans, moreover, show that even restricted lesions of the cortex or thalamus alone can result in an inability to stay awake normally (25, 26). But no inhibitory effects of the cortex in sleep have yet been uncovered.

In sexual behavior it has been found that lesions of the new cortex may interfere directly with the arousal of sexual behavior (9, 11). Large lesions are much more effective than small lesions, as you might expect. Furthermore, cortical damage is much more serious in male animals than in females and is much more important in the sexual behavior of primates than it is in the case of lower mammals. On the other hand, in connection with studies of the cortex in emotions, it has been found that lesions of the amygdala and transitional cortex of the midline can lead to heightened sexuality in cats and monkeys (7, 40). So it looks as though the cortex may exert both excitatory and inhibitory influences in sexual motivation.

Evidence from other types of motivated behavior is only fragmentary, but it fits into the same general picture. In

the case of hunger, it has been reported that certain lesions of the frontal lobes will lead to exaggerated eating behavior (41, 55). Hyperactivity may follow similar frontal lobe lesions and is particularly marked after damage to the orbital surface of the frontal lobe (56). The frontal areas may also be involved in what might be called pain avoidance. Clinical studies of man show that lobotomies may be used for the relief of intractable pain (29). The curious thing about these cases is that they still report the same amount of pain after operation but they say that it no longer bothers them. Presumably the frontal cortex normally plays an excitatory role in the motivation to avoid pain.

In all the cases cited so far, the anatomical and physiological evidence available suggests strongly that the main influence of the cortex and thalamus in motivation is mediated by the hypothalamus. But we do not yet have direct proof of this point and need experiments to check it.

Interaction of factors. Up to now, we have treated the various factors that can operate in the control of motivated behavior singly. However, one of the main points of the theory proposed here is that the various factors operate together in the control of motivation. Presumably this interaction of factors occurs in the hypothalamus and takes the form of the "addition" of all excitatory influences and the "subtraction" of all inhibitory influences. Some experimental evidence bears directly on this point.

In the case of sexual behavior, for example, it is clear that excitatory influences of the cortex and hormones are additive. After sexual motivation is eliminated by cortical damage it may be restored by the administration of large doses of sex hormones (10). Since the hypothalamus is the site of action of the sex hormones, it seems likely that

it is also the site of interaction of the influences of the hormones and cortex.

In a similar way, it looks as though the contributions of sensory stimulation and sex hormones add in the hypothalamus. Neither hormones nor stimulation alone is sufficient to elicit sexual reactions in most mammals, but the right combination of the two will. Still another example of the addition of excitatory influences is seen in the study of the sexual behavior of the male rabbit. In this case neither destruction of the olfactory bulbs nor decortication will eliminate mating behavior, but a combination of the two operations will (21).

It is very important to know whether excitatory, and perhaps also inhibitory, influences in other kinds of motivation have the same sort of additive properties as in sexual behavior. Indirect evidence suggests they do, but direct experiments of the sort described here are needed to check the possibility.

Most encouraging in this connection is that students of instinctive behavior in inframammalian vertebrates and invertebrates have presented considerable evidence showing that sensory, chemical, and neural influences contribute jointly to the arousal of many kinds of motivated behavior (60). For example, in a number of cases it has been shown that the threshold for arousing behavior by various stimuli is lowered considerably by appropriate changes in the internal environment. In fact, in the extreme case, when internal changes are maximal, the behavior may occur in the absence of any obvious stimulation. Presumably in these cases, as in the examples of mammalian motivation, chemical and neural influences contribute to the arousal of some central response mechanism in an additive way.

The role of learning. It is obvious to every student of mammalian motivation

that learning and experience may play extremely important roles in the regulation of motivated behavior. What does this mean in terms of the present physiological theory? Unfortunately, we cannot specify the mechanisms through which learning enters into the control of motivation because we are ignorant of the basic physiology of learning. But we can make some helpful inferences.

The basic hypothesis in the present theoretical framework is that learning contributes to hypothalamic activity along with influences from unlearned afferent impulses, internal changes, and cortical activity. In the case of sexual behavior we know that many animals learn to be aroused sexually by stimuli which were not previously adequate. Further, we know that in such experienced animals it is difficult to reduce sexual motivation by eliminating avenues of sensory stimulation, presumably because the extra excitatory effects produced by learned stimuli contribute to hypothalamic activity along with the impulses from unlearned stimuli. Along the same lines, it is known that sex hormones are relatively unimportant in man and in certain of the sub-human primates that have learned to be aroused by a wide variety of stimuli (12). Again, this may mean that the excitatory effects from the learned stimuli have added enough to the effects of unlearned stimuli to make it possible to dispense with the contribution of the sex hormones in arousing hypothalamic activity.

The evidence available on learning in other types of motivation fits in with this general theoretical picture, but direct physiological experiments have not yet carried us beyond the stage of inference. We know, for example, that vitamin-deficient rats can learn to show motivated behavior in response to cer-

tain flavors that have been associated with the vitamin in the past (34, 58). In fact, for a short while they will even pass up food containing the vitamin to eat vitamin-deficient food containing the flavor. Again, it looks as though flavor has become empowered by a process of learning to contribute to the excitability of the neural centers controlling motivation.

LIMITATIONS OF THE THEORY

Like any theoretical approach, the physiological mechanism proposed here has many limitations. Fortunately none of them need be too serious as long as it is recognized that the theory is set up as a general guide for experiments and a framework for further theorizing. Obviously the theory is going to have to be changed and improved many times before it is free of limitations. In this spirit it might be said that the limitations of the theory are not much more than those aspects of motivation which need research the most. But whether we label them limitations or urgent areas of research, they deserve explicit attention.

The concept of "center." Throughout this discussion the terms "neural center" and "hypothalamic center" have been used. "Center" is a useful and convenient term, but it is also a dangerous one, for it may carry with it the implication of strict localization of function within isolated anatomical entities. Actually this implication is not intended, for it is recognized that localization is a relative matter and that no neural mechanism operates in isolation. Furthermore, it is also possible that there may be no discoverable localization of the neural mechanisms governing some types of motivated behavior. The theory simply states at the moment that the best general hypothesis is that some degree of localization of the mechanisms

controlling motivation can be found in the hypothalamus.

Execution of motivated behavior. No attempt has been made in this discussion to describe the details of the efferent pathways or effector mechanisms responsible for the execution of motivated behavior. Discussion of the pathways has been omitted because we know very little about them. About all we can do at present is to guess, from anatomical and physiological studies of hypothalamic function, that the hypothalamus exerts some kind of "priming" effect on effector pathways controlled by other parts of the nervous system. Perhaps after the relationship of the hypothalamus to motivated behavior has been more firmly established we can profitably turn to the question of how the hypothalamus does its work.

A second aspect of the execution of motivated behavior has been omitted for the sake of brevity. We all recognize that an animal with certain kinds of cortical lesions, or deprived of certain sensory capacities, may be handicapped in executing motivated behavior quite aside from any effects these operations may have on the arousal of motivation. Fortunately most investigators have been aware of this problem and have taken pains to distinguish these two effects, focusing their attention mainly on the arousal of motivation. Some day, however, this theory should address the question of what neural mechanisms govern the execution of motivated behavior.

General nature of the mechanism. For theoretical purposes it has been assumed that essentially the same mechanism controls all types of motivated behavior. Obviously this is not likely to be the case, nor is it an essential assumption. In some types of motivation only parts of this mechanism may be involved, or factors not included in the present scheme may operate. For

example, in some cases the hypothalamus may not be involved at all, or it may turn out that there are no inhibitory centers at work, or that internal chemical factors do not contribute significantly. There is no reason why we should not be prepared for these eventualities. But until specific experimental evidence to the contrary is forthcoming, the general mechanism proposed here still remains as the best working hypothesis for any particular type of biological motivation.

Inadequacy of behavioral measures. To a large degree the present discussion is based upon measures of consummatory behavior. We all know that the various measures of motivation are not always in good agreement, so there is good possibility that what we say about consummatory behavior may not apply to motivation measured by other methods. In fact, Miller, Bailey, and Stevenson (46) have recently shown that whereas rats with hypothalamic lesions overeat in the free-feeding situation, they do not show a high degree of motivation when required to overcome some barrier to obtain food.

Confining the present discussion mainly to consummatory behavior is clearly a weakness. But the logic behind this limited approach is to work out the physiological mechanisms in the simplest case first, and then to see how they must be revised to fit the more complicated cases.

Complex motivation. It can also be argued, of course, that the present theory is confined to the simple, biological motives. Again, it seems eminently advisable to keep the theory relatively narrow in scope until it is developed well enough to permit attack on the more complicated, learned motives.

Comparative approach. No attempt has been made here to make it explicit how the proposed theory applies to organisms representative of different phy-

logenetetic levels. There are many obvious advantages to the comparative approach, but unfortunately, except for the case of sexual motivation, the information we have on different species is too scattered to be useful. Judging from what we have learned from the comparative study of sexual motivation, however, we can expect the various factors governing other types of motivation to contribute somewhat differently in animals at different phylogenetic levels. Certainly learning should be more important in primates than in subprimates, and the contributions of the cortex and thalamus should be greater. Much will be gained if future research in motivation follows the excellent example set in the study of sexual behavior and provides the much needed comparative data.

ADVANTAGES OF THE THEORY

On the assumption that none of these limitations of the theory are critical, it is appropriate to ask: What is gained by proposing an explicit theory of the physiological mechanisms underlying motivated behavior? There are many positive answers to this question, and we can list some of them briefly.

Simplification of the problem. One of the main advantages of the theoretical mechanism proposed here is that it brings together, into one general framework, a number of different kinds of motivation that have been studied separately in the past. Certainly the theory encompasses the basic facts available on sex, hunger, specific hunger, thirst, sleep, and emotion. And it may also be able to handle the facts of pain avoidance, hoarding, nesting, maternal behavior, and other types of so-called instinctive behavior. As you have seen, one of the benefits deriving from this kind of simplification of the problem of motivation is the possibility of speeding

up progress by applying what has been learned about physiological mechanisms from the study of one kind of motivation to the study of other kinds of motivation. Not only that, but the assumption that the hypothalamus is central in the control of all types of motivation may make it easier to explain the various types of interaction among motivations that have shown up in many studies of behavior.

Multifactor approach. Another advantage of the present theory is that it gives strong emphasis to the view that motivation is under multifactor control. Single-factor theories, so prevalent since the days of Cannon, can only lead to useless controversies over which factor is the "right" one and must always be guilty of omission in trying to account for the control of motivation. Of course, it must be stressed that the aim of the multifactor approach is not simply to list the many possible factors operating in motivation, but rather to get down to the concrete experimental task of determining the relevant factors which control motivation and the relative contribution of each.

Satiation of motivation. Unlike most previous theories of motivation, the mechanism proposed here attempts to account for the satiation of motivation as well as its arousal. In terms of the present theory satiation is determined by the reduction of activity in the main excitatory centers of the hypothalamus. More specifically, it looks as though the inhibitory centers of the hypothalamus may constitute a separate "satiation mechanism" which is the most important influence in the reduction of the activity of the excitatory centers. The possibility is an intriguing one, and it can be directly explored by experiment.

Peripheral and central control. In the past the study of motivation has been hampered by the controversy over whether behavior is centrally or periph-

erally controlled. The controversy is nonsense. The only meaningful experimental problem is to determine how the central and peripheral, or sensory, factors operate together in the control of behavior. It is this problem which the present theory addresses directly, and this is one of its greatest strengths.

Learned and innate control. The present theory avoids another knotty controversy by directly addressing experimental problems. Much time has been lost in psychology, and particularly in the study of motivation, in arguments over whether behavior is primarily innate or instinctive or whether it is primarily learned or acquired. The answer is obviously that it is both, and again the only meaningful experimental problem is to determine the relative contribution of each type of control. As far as the mechanism proposed here is concerned, both innate and learned factors make their contributions to the control of the same hypothalamic centers. There is still much work needed to determine the details of the mechanisms of operation, particularly of the learned factors, but some headway has been made and the problem is clearly set.

Explicit nature of the theory. Finally, a number of advantages derives simply from having an explicit statement of an up-to-date, physiological theory of motivation. In the first place, an explicit theory can serve as a convenient framework within which to organize the physiological facts we already have at our disposal. Second, the systematic organization of the facts sharply points up many of the gaps in our knowledge and suggests direct experiments that should be done in the investigation of motivated behavior. Third, an up-to-date, systematic theory provides a useful and reasonably clear conceptualization of motivation for psychologists working in other areas of research.

SUMMARY AND CONCLUSIONS

A physiological theory of motivated behavior is presented. The basic assumption in this theory is that the amount of motivated behavior is a function of the amount of activity in certain excitatory centers of the hypothalamus. The level of activity of the critical hypothalamic centers, in turn, is governed by the operation of four factors.

1. Inhibitory centers in the hypothalamus directly depress the activity of the excitatory centers and may be responsible for the production of satiation.

2. Sensory stimuli set up afferent impulses which naturally contribute to the excitability of the hypothalamus or come to do so through a process of learning.

3. Changes in the internal environment exert both excitatory and inhibitory effects on the hypothalamus.

4. Cortical and thalamic influences increase and decrease the excitability of hypothalamic centers.

Detailed experimental evidence is brought forward to show how these various factors operate in the management of different kinds of motivated behavior. The over-all scheme is shown diagrammatically in Fig. 1.

Out of consideration of this evidence a number of hypotheses are generated to fill in the gaps in experimental knowledge. All these hypotheses are experimentally testable. The ones of major importance can be given here as a summary of what the theory states and a partial list of the experiments it suggests.

1. There are different centers in the hypothalamus responsible for the control of different kinds of basic motivation.

2. In each case of motivation, there is one main excitatory center and one inhibitory center which operates to de-

press the activity of the excitatory center.

There is already much experimental evidence supporting these two general hypotheses, but it is not certain that they apply fully to all types of basic biological motivation. The hypotheses should be checked further by determining whether changes in all types of motivation can be produced by local hypothalamic lesions and whether both increases and decreases in motivation can always be produced.

3. The activity of hypothalamic centers is, in part, controlled by the excitatory effects of afferent impulses generated by internal and external stimuli.

4. Different stimuli contribute different relative amounts to hypothalamic activity but no one avenue of sensory stimulation is indispensable.

5. It is the sum total of afferent impulses arriving at the hypothalamus that determines the level of excitability and, therefore, the amount of motivation.

The neuroanatomical and neurophysiological evidence shows that the hypothalamus is richly supplied with afferents coming directly and indirectly from all the sense organs (Fig. 2). The behavioral evidence, furthermore, strongly suggests that motivation is never controlled, in mammals at least, by one sensory system, but rather is the combination of contributions of several sensory systems. Sensory control and sensory deprivation experiments are needed to check this point in the case of most kinds of biological motivation, particularly hunger, thirst, and specific hungers.

6. A variety of kinds of physical and chemical changes in the internal environment influences the excitability of hypothalamic centers and, therefore, contributes to the control of motivation.

The evidence shows that the hypothalamus is the most richly vascularized region of the central nervous system and

is most directly under the influence of the cerebrospinal fluid. Furthermore, it is clear that changes in the internal environment produced by temperature of the blood, osmotic pressure, hormones, and a variety of other chemicals are important in motivation and most likely operate through their influence on the hypothalamus. Direct studies are still needed in many cases, however, to show that the particular change that is important in motivation actually does operate through the hypothalamus and vice versa.

7. The cerebral cortex and thalamus are directly important in the temporal and spatial organization of motivated behavior.

8. Different parts of the cortex and thalamus also operate selectively in the control of motivation by exerting excitatory or inhibitory influences on the hypothalamus.

Tests of these hypotheses can be carried out by total decortication, partial cortical ablations, and local thalamic lesions. It should be especially instructive to see what effects cortical and thalamic lesions have after significant changes in motivation have been produced by hypothalamic lesions.

9. Learning contributes along with other factors to the control of motivation, probably through direct influence on the hypothalamus.

10. The relative contribution of learning should increase in animals higher and higher on the phylogenetic scale.

A whole series of experiments is needed here. Particularly, there should be comparisons of naive and experienced animals to determine the relative effects of sensory deprivation, cortical and thalamic damage, and hypothalamic lesions. Presumably animals that have learned to be aroused to motivated behavior by previously inadequate stimuli should require more sensory deprivation

but less cortical and thalamic damage than naive animals before motivation is significantly impaired.

11. The various factors controlling motivation combine their influences at the hypothalamus by the addition of all excitatory influences and the subtraction of all inhibitory influences.

Some experiments have already been done in the study of sexual motivation to show that motivation reduced by the elimination of one factor (cortical lesions) can be restored by increasing the contribution of other factors (hormone therapy). Many combinations of this kind of experiment should be carried out with different kinds of motivated behavior.

A number of the limitations and some of the advantages of the present theoretical approach to the physiology of motivation are discussed.

REFERENCES

- ADOLPH, E. F. The internal environment and behavior. Part III. Water content. *Amer. J. Psychiat.*, 1941, **97**, 1365-1373.
- ADOLPH, E. F. Thirst and its inhibition in the stomach. *Amer. J. Physiol.*, 1950, **161**, 374-386.
- ANAND, B. K., & BROBECK, J. R. Hypothalamic control of food intake in rats and cats. *Yale J. Biol. Med.*, 1951, **24**, 123-140.
- ANAND, B. K., & BROBECK, J. R. Localization of a "feeding center" in the hypothalamus of the rat. *Proc. Soc. exp. Biol. Med.*, 1951, **77**, 323-324.
- BARD, P. Central nervous mechanisms for emotional behavior patterns in animals. *Res. Publ. Ass. nerv. ment. Dis.*, 1939, **19**, 190-218.
- BARD, P. The hypothalamus and sexual behavior. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, **20**, 551-579.
- BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in the expression of rage with special reference to the suppression of angry behavior. *Res. Publ. Ass. nerv. ment. Dis.*, 1947, **27**, 362-404.
- BEACH, F. A. Analysis of factors involved in the arousal, maintenance and manifestation of sexual excitement in male animals. *Psychosom. Med.*, 1942, **4**, 173-198.
- BEACH, F. A. Central nervous mechanisms involved in the reproductive behavior of vertebrates. *Psychol. Bull.*, 1942, **39**, 200-206.
- BEACH, F. A. Relative effect of androgen upon the mating behavior of male rats subjected to forebrain injury or castration. *J. exp. Zool.*, 1944, **97**, 249-295.
- BEACH, F. A. A review of physiological and psychological studies of sexual behavior in mammals. *Physiol. Rev.*, 1947, **27**, 240-307.
- BEACH, F. A. Evolutionary changes in the physiological control of mating behavior in mammals. *Psychol. Rev.*, 1947, **54**, 297-315.
- BELLOWS, R. T. Time factors in water drinking in dogs. *Amer. J. Physiol.*, 1939, **125**, 87-97.
- BREMER, F. Étude oscillographique des activités sensorielles du cortex cérébral. *C. r. Soc. Biol.*, 1937, **124**, 842-846.
- BROBECK, J. R. Regulation of energy exchange. In J. F. Fulton (Ed.), *A textbook of physiology*. Philadelphia: Saunders, 1950. Pp. 1069-1090.
- BROBECK, J. R., TEPPERMAN, J., & LONG, C. N. H. Experimental hypothalamic hyperphagia in the albino rat. *Yale J. Biol. Med.*, 1943, **15**, 831-853.
- BROMILEY, R. B., & BARD, P. A study of the effect of estrin on the responses to genital stimulation shown by decapitate and decerebrate female cats. *Amer. J. Physiol.*, 1940, **129**, 318-319.
- BROOKHART, J. M., & DEY, F. L. Reduction of sexual behavior in male guinea pigs by hypothalamic lesions. *Amer. J. Physiol.*, 1941, **133**, 551-554.
- BROOKHART, J. M., DEY, F. L., & RANSON, S. W. Failure of ovarian hormones to cause mating reactions in spayed guinea pigs with hypothalamic lesions. *Proc. Soc. exp. Biol. Med.*, 1940, **44**, 61-64.
- BROOKHART, J. M., DEY, F. L., & RANSON, S. W. The abolition of mating behavior by hypothalamic lesions in guinea pigs. *Endocrinology*, 1941, **28**, 561-565.
- BROOKS, C. M. The role of the cerebral cortex and of various sense organs in the excitation and execution of mating activity in the rabbit. *Amer. J. Physiol.*, 1937, **120**, 544-553.
- BROOKS, C. M. Appetite and obesity. *N. Z. med. J.*, 1947, **46**, 243-254.
- CANNON, W. B. Hunger and thirst. In C. Murchison (Ed.), *A handbook of*

- general experimental psychology. Worcester, Mass.: Clark Univer. Press, 1934. Pp. 247-263.
24. CRAIGIE, E. H. Measurements of vascularity in some hypothalamic nuclei of the albino rat. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, 20, 310-319.
 25. DAVISON, C., & DEMUTH, E. L. Disturbances in sleep mechanism: a clinico-pathologic study. I. Lesions at the cortical level. *Arch. Neurol. Psychiat.*, Chicago, 1945, 53, 399-406.
 26. DAVISON, C., & DEMUTH, E. L. Disturbances in sleep mechanism: a clinico-pathologic study. II. Lesions at the corticodiencephalic level. *Arch. Neurol. Psychiat.*, Chicago, 1945, 54, 241-255.
 27. DELGADO, J. M. R., & ANAND, B. K. Increase of food intake induced by electrical stimulation of the lateral hypothalamus. *Amer. J. Physiol.*, 1953, 172, 162-168.
 28. DEMPSEY, E. W., & RIOCH, D. McK. The localization in the brain stem of the oestrous responses of the female guinea pig. *J. Neurophysiol.*, 1939, 2, 9-18.
 29. FREEMAN, W., & WATTS, J. W. *Psychosurgery*. (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1950.
 30. GELLHORN, E. *Autonomic regulations*. New York: Interscience, 1943.
 31. GILMAN, A. The relation between blood osmotic pressure, fluid distribution and voluntary water intake. *Amer. J. Physiol.*, 1937, 120, 323-328.
 32. GROSSMAN, M. I., CUMMINS, G. M., & IVY, A. C. The effect of insulin on food intake after vagotomy and sympathectomy. *Amer. J. Physiol.*, 1947, 149, 100-102.
 33. GROSSMAN, M. I., & STEIN, I. F. Vagotomy and the hunger producing action of insulin in man. *J. appl. Physiol.*, 1948, 1, 263-269.
 34. HARRIS, L. J., CLAY, J., HARGREAVES, F. J., & WARD, A. Appetite and choice of diet. The ability of the Vitamin B deficient rat to discriminate between diets containing and lacking the vitamin. *Proc. roy. Soc.*, 1933, 113, 161-190.
 35. HETHERINGTON, A. W., & RANSON, S. W. The spontaneous activity and food intake of rats with hypothalamic lesions. *Amer. J. Physiol.*, 1942, 136, 609-617.
 36. INGRAM, W. R. Nuclear organization and chief connections of the primate hypothalamus. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, 20, 195-244.
 37. JANOWITZ, H. D., & GROSSMAN, M. I. Some factors affecting the food intake of normal dogs and dogs with esophagostomy and gastric fistula. *Amer. J. Physiol.*, 1949, 159, 143-148.
 38. KLEITMAN, N. *Sleep and wakefulness*. Chicago: Univer. of Chicago Press, 1939.
 39. KLEITMAN, N., & CAMILLE, N. Studies on the physiology of sleep. VI. Behavior of decorticated dogs. *Amer. J. Physiol.*, 1932, 100, 474-480.
 40. KLÜVER, H., & BUCY, P. C. Preliminary analysis of functions of the temporal lobes in monkeys. *Arch. Neurol. Psychiat.*, Chicago, 1939, 42, 979-1000.
 41. LANGWORTHY, O. R., & RICHTER, C. P. Increased spontaneous activity produced by frontal lobe lesions in cats. *Amer. J. Physiol.*, 1939, 126, 158-161.
 42. LASHLEY, K. S. Experimental analysis of instinctive behavior. *Psychol. Rev.*, 1938, 45, 445-471.
 43. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 473-516.
 44. MAGOUN, H. W., HARRISON, F., BROBECK, J. R., & RANSON, S. W. Activation of heat loss mechanisms by local heating of the brain. *J. Neurophysiol.*, 1938, 1, 101-114.
 45. MAYER, J., VITALE, J. J., & BATES, M. W. Mechanism of the regulation of food intake. *Nature, London*, 1951, 167, 562-563.
 46. MILLER, N. E., BAILEY, C. J., & STEVENSON, J. A. F. Decreased 'hunger' but increased food intake resulting from hypothalamic lesions. *Science*, 1950, 112, 256-259.
 47. MORGAN, C. T. *Physiological psychology*. (1st Ed.) New York: McGraw-Hill, 1943.
 48. MORGAN, C. T., & MORGAN, J. D. Studies in hunger. 1. The effects of insulin upon the rat's rate of eating. *J. genet. Psychol.*, 1940, 56, 137-147.
 49. NAUTA, W. J. H. Hypothalamic regulation of sleep in rats; an experimental study. *J. Neurophysiol.*, 1946, 9, 285-316.
 50. RANSON, S. W. Somnolence caused by hypothalamic lesions in the monkey. *Arch. Neurol. Psychiat.*, 1939, 41, 1-23.
 51. RANSON, S. W. Regulation of body temperature. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, 20, 342-399.
 52. RANSON, S. W., KABAT, H., & MAGOUN, H. W. Autonomic responses to electrical stimulation of hypothalamus, pre-

- optic region and septum. *Arch. Neurol. Psychiat., Chicago*, 1935, 33, 467-477.
53. RANSTRÖM, S. *The hypothalamus and sleep regulation*. Uppsala: Almqvist and Wiksells, 1947.
54. RICHTER, C. P. Total self regulatory functions in animals and human beings. *Harvey Lect.*, 1942-43, 38, 63-103.
55. RICHTER, C. P., & HAWKES, C. D. Increased spontaneous activity and food intake produced in rats by removal of the frontal poles of the brain. *J. Neurol. Psychiat.*, 1939, 2, 231-242.
56. RUCH, T. C., & SHENKIN, H. A. The relation of area 13 of the orbital surface of the frontal lobe to hyperactivity and hyperphagia in monkeys. *J. Neurophysiol.*, 1943, 6, 349-360.
57. SANGSTER, W., GROSSMAN, M. I., & IVY, A. C. Effect of d-amphetamine on gastric hunger contractions and food intake in the dog. *Amer. J. Physiol.*, 1948, 153, 259-263.
58. SCOTT, E. M., & VERNEY, E. L. Self selection of diet. VI. The nature of appetites for B vitamins. *J. Nutrit.*, 1947, 34, 471-480.
59. SOULAIRAC, A. La physiologie d'un comportement: L'appétit glucidique et sa régulation neuro-endocrinienne chez les rongeurs. *Bull. Biol.*, 1947, 81, 1-160.
60. TINBERGEN, N. *The study of instinct*. London: Oxford Univer. Press, 1951.
61. VERNEY, E. B. The antidiuretic hormone and the factors which determine its release. *Proc. roy. Soc., London*, 1947, 135, 24-106.
62. WHEATLEY, M. D. The hypothalamus and affective behavior in cats. *Arch. Neurol. Psychiat.*, 1944, 52, 296-316.

(Received February 26, 1953)

THE S-R REINFORCEMENT THEORY OF EXTINCTION

HENRY GLEITMAN, JACK NACHMIAS,¹ AND ULRIC NEISSER²

Swarthmore College

Stimulus-response reinforcement theory as formulated by Hull (10), the most highly developed of current learning theories, has been the center of much debate and controversy. Its view of reinforcement has been challenged by the latent-learning studies (1, 32), its conception of the response has been attacked by place-learning experiments (27, 33), and its analysis of discrimination learning has been repeatedly questioned, both by the adherents of noncontinuity theories (16, 18), and more recently by other writers (28). Comparatively little attention, however, has been paid to its theory of extinction.

This omission is regrettable, in view of the fact that extinction constitutes a strategic area for any learning theory. In the first place, it represents an important phenomenon which every theory must at least attempt to explain—adaptive behavior presupposes not only the acquisition of appropriate new responses, but also the abandonment of inappropriate old ones. Furthermore, theoretical interpretations of extinction play an important part in the explanation of other phenomena; thus most S-R theorists consider discrimination learning to be the result of an interaction between excitatory and inhibitory tendencies.

This paper³ will examine the S-R reinforcement theory of extinction, and will try to show that it suffers

from some serious shortcomings. Specifically, we believe that Hull's theory of extinction does not fit all the experimental facts, involves certain conceptual difficulties, and generates some paradoxical predictions.

HULL'S THEORY OF EXTINCTION

Following Hilgard and Marquis (7), theories of extinction can be grouped into two general categories: interference and adaptation. Interference theories, such as Guthrie's (6) and Wendt's (34), assert that extinction is due to the association of interfering responses to the conditioned stimulus. Adaptation theories, such as Hull's theory of reactive inhibition, assume that extinction is caused by an inhibitory factor generated by the repeated elicitation of the response. This inhibitory factor—believed to be analogous to fatigue—is said to act against the further evocation of the response, and is usually thought to dissipate with time.

Razran (26) and Hilgard and Marquis have shown that neither of these theories by itself provides an adequate explanation of the phenomena of extinction. Interference theories fail to indicate how the interfering responses arise in the first place. They do not account for spontaneous recovery, although recent attempts in that direction have been made by Liberman (19, 20). They are further challenged by certain facts concerning the rates of conditioning and extinction. If extinction were but a manifestation of the conditioning of interfering responses, then any factor that facilitates conditioning should like-

¹ Now at Harvard University.

² Now at Massachusetts Institute of Technology.

³ We wish to express our appreciation to Dr. Edward Walker for his helpful suggestions and criticisms.

wise accelerate extinction. In actual fact, stimulants increase the rate of conditioning and retard extinction while depressants retard conditioning but accelerate extinction. The negative correlation usually found between rates of conditioning and rate of extinction likewise argues against an interference theory (7, p. 119).

An adaptation theory alone is also inadequate. It fails to account for the fact that spontaneous recovery is usually incomplete, and that repeated extinction sessions eventually lead to a total lack of recovery. It does not explain the stimulus generalization of extinction effects nor the phenomenon of disinhibition.

Hull's theory of extinction (10),⁴ like that presented by Miller and Dollard (24), utilizes both interference and adaptation concepts and thus has a considerably expanded scope. It first postulates the operation of an inhibitory factor, reactive inhibition or I_R , which tends to counteract the further occurrence of the response. This factor is assumed to result from the elicitation of the response itself, to vary with the effort involved in the performance of that response, and to decay with time. On this basis, Hull deduces a variety of phenomena such as spontaneous recovery, the superiority of distributed over massed practice, and reminiscence.

In addition to mere effector inhibition (I_R), Hull also postulates that extinction involves the production of a habit, conditioned inhibition or sI_R —a habit of *not* responding. Its origin is explained as follows:

⁴ Recently, Hull's systematic formulations have been revised and elaborated in *Essentials of Behavior* (11), and in *A Behavior System* (12). Since we feel that these more recent publications have left the theory of extinction essentially unaltered, we shall base our discussion primarily upon the more familiar *Principles of Behavior*.

... the after-effects of response evocation in the aggregate constitute a negative drive strongly akin to tissue injury or "pain." If this is the case, we should expect that the cessation of the "noxious" stimulation in question or the reduction in the inhibitory substance, or both, would constitute a reinforcing state of affairs. The response process which would be most closely associated with such a reinforcing state of affairs would obviously be the cessation of the activity itself. In accordance with the "law of reinforcement" ... this cessation of activity would be conditioned to any afferent stimulus impulse, or stimulus traces, which chanced to be present at the time the need decrement occurred. Consequently there would arise the somewhat paradoxical phenomenon of a negative habit, i.e., a habit of *not* doing something" (10, p. 282).

Being a habit, sI_R is tied to a stimulus, and presumably does not dissipate with time. It can thus be invoked to explain the generalization of extinction along stimulus dimensions, disinhibition, and the incompleteness of spontaneous recovery.

Both inhibitory factors, sI_R and I_R , contribute to the extinction process by summing to make up an inhibitory aggregate \dot{I}_R , which is subtracted from reaction potential, sE_R , to yield effective reaction potential, $s\bar{E}_R$. This relation is expressed by the following equations:

$$\dot{I}_R = sI_R + I_R, \quad s\bar{E}_R = sE_R - \dot{I}_R$$

It is important to note that, according to Hull, both sI_R and I_R are produced during rewarded as well as during unrewarded trials; the rise of the learning curve during conditioning only means that each response leads to a greater increment of reaction potential than of the inhibitory aggregate.

On the surface, Hull's theory of extinction seems to account for many of the facts with considerable elegance. Nevertheless, we believe that his conception of the extinctive process is beset by serious problems. We shall

discuss these problems under three headings, in what we feel is the order of increasing importance: (a) empirical difficulties, (b) conceptual difficulties, and (c) some paradoxical predictions generated by the theory.

EMPIRICAL DIFFICULTIES

According to Hull's theory of extinction, the elimination of a response presupposes the performance of the response to be eliminated, or at least the performance of another response from which extinction effects can generalize. For both reactive and conditioned inhibition depend upon response performance, the former directly, and the latter indirectly through its dependence on I_R reduction. There are some experimental findings, however, which at least suggest that the performance of an activity is not a necessary condition for its extinction.

1. *Subzero extinction.* Evidence for such a possibility comes first from the phenomenon of "subzero extinction," demonstrated in classical conditioning. Pavlov (25) showed that when a conditioned response has been extinguished to the point of nonelicitation, further unreinforced presentations of the conditioned stimulus will nevertheless serve to strengthen extinction, as measured by a decrease in spontaneous recovery. Similar results were obtained by Brogden, Lipman, and Culler (2).

It might be argued that these effects are the result of the extinction of covert, implicit responses, which were elicited even when the overt ones were absent. Such an interpretation is consonant with the findings of Brogden, Lipman, and Culler (2) that slight forelimb movements did persist into the subzero extinction trials. This implies that crucial implicit responses or r_g 's survived the elimina-

tion of the overt, "parent" responses, and that they are thus more resistant to extinction than are the latter.

2. *Latent extinction.* Further evidence comes from studies reporting an effect which might be called "latent extinction" by analogy with the phenomenon of latent learning. There have recently been three experiments in this area.

Seward and Levy (29) trained rats to run a straight alley with food on the goal platform. Subsequently, the animals were extinguished in two different ways: The experimental group was detained on the now empty goal platform both before and between extinction trials, whereas the control group spent equivalent periods on a *neutral* empty platform. The experimental group reached the extinction criterion in significantly fewer trials, and ran more slowly than the control group. The effect of previous detention on the empty goal platform appeared even on the very first extinction trial: compared with training there was a significant decrease in running time for the experimental animals after such treatment, whereas the corresponding decrease for the control animals was not significant. This suggests that an instrumental response can be extinguished without being elicited.

Bugelski, Coyer, and Rogers (3) took issue with the experimental design employed by Seward and Levy, pointing out that their experimental and control animals were detained on different platforms even during the extinction procedure (between trials), so that the test situation was not identical for both groups. (We don't entirely understand this objection, since Seward and Levy had already found a significant difference between the running times of their two groups on the first test trial.) Upon repeat-

ing the experiment, Bugelski, Coyer, and Rogers failed to obtain any evidence of latent extinction. There is some doubt, however, whether the repetition really duplicated the conditions of the earlier experiment, since even the control animals used by Seward and Levy gave up running sooner than did those employed in the replication. Bugelski, Coyer, and Rogers suggest that this difference may be due to an age factor; the rats used in their experiment were younger and may have been more active.

Latent extinction, however, was also obtained in an experiment by Deese (4). He trained rats to run to one side of a U maze, and afterwards was able to extinguish the correct choice response by merely placing the animals in the goal box without food. Animals who were subjected to this nonresponse extinction procedure made a smaller proportion of correct choices when again run in the maze than did control animals who were not permitted to "inspect" the empty goal box. Thus, again some extinction occurred without the prior performance of the response to be extinguished, and therefore did not seem to depend upon response-produced inhibition. In fact, nonresponse extinction was just about as effective as response extinction in producing the abandonment of the correct response. Comparing the results on four ordinary, nonreinforced response trials that were preceded in one group by four nonresponse extinction trials, and in the other group by four response extinction trials, we find hardly any difference. In other words, being placed in an empty goal box four times seems just as effective in reducing the likelihood of running subsequently as actually having run there on four occasions.

This conclusion is based upon the

two groups which had eight consecutive extinction trials on the same day. The effect is even more striking with groups which were given a 24-hr. rest interval between the first four and second four extinction trials. In these groups, four exposures to the empty goal box led to a greater decrement in performance, measured on the second day, than four nonreinforced runs. In other words, response extinction showed the effects of spontaneous recovery, while "nonresponse" extinction did not.

It might be asserted that these results can be explained in a manner similar to that in which Spence (31) and others (22) have attempted to deal with latent learning; that is, by reference to fractional anticipatory goal responses— r_g 's—to whose sensory consequences the turning or running responses had become conditioned during training, and which have become extinguished during the latent-extinction period. Such an explanation does not seem plausible.

Since the r_g is an implicit response, it presumably requires very little effort. It follows that many trials should be required for its extinction. As we have seen, this deduction is confirmed by the data from subzero extinction, at least if these are to be explained by the supposed extinction of implicit responses. Yet Deese's animals, given only four trials in the empty goal box, nevertheless showed extinction effects equal in magnitude to those of animals which were required to actually run to the goal box. If his results are also ascribed to the ubiquitous r_g , this now possesses somewhat contradictory properties.

The preceding discussion has not exhausted the empirical difficulties encountered by Hull's theory of extinction. To give merely two examples, the roles played by effort and

by spacing in learning and extinction are by no means clear. For reasons of brevity, we have confined ourselves to the discussion of what we consider the most central empirical question: Is performance a necessary condition for the extinction of a response?

CONCEPTUAL DIFFICULTIES

The S-R reinforcement theory of extinction has shortcomings more serious than the empirical problems discussed above. These shortcomings become apparent when we try to discover how the theory's central constructs are conceptualized. We will confine ourselves here to a discussion of the *conditioned inhibition* construct, which seems to pose the most serious problems. In so doing, however, we do not wish to minimize the difficulties involved in the notion of *reactive inhibition*. The latter is usually discussed (10, 30) as if it were a result of proprioceptive stimulation from the specific effectors involved in the response, yet the learned response is often defined, not in terms of specific effectors, but in the broader terms demanded by the results of place learning (27) and response generalization (18) experiments. Thus, Miller and Dollard define the response as "any activity within the individual which can become functionally connected with an antecedent event through learning" (24, p. 59). A related problem arises from the results of Gustafson and Irion (21, p. 174) and Kimble (14), who have shown a clear reminiscence effect in bilateral transfer. They point out that if reminiscence is to be ascribed to the dissipation of I_R , then I_R must inhibit more than a particular, specific response. We refrain from extended discussion of these matters, however, because we feel that the difficulties

can probably be overcome by clarification of the definitions involved.

The concept of reactive inhibition requires more thorough consideration. Originally sI_R —the "habit of not responding"—was proposed by Hull to account for the more stable aspects of extinction, and for the stimulus generalization of extinction effects. The act of not responding, from here on referred to as the "not-response," is connected to the stimulus situation by the reinforcing effects of I_R reduction. The not-response is treated as formally equivalent to an ordinary response, so that sI_R is really an S-not-R bond; thus the laws of habit formation are widened to include extinction.

In order to understand what might be meant by a "habit of not responding," we must first be clear on the meaning of the not-response itself. This consideration is all the more appropriate since the postulation of not-responses that have the same status as ordinary responses has become increasingly widespread in S-R reinforcement theory. For instance, Dollard and Miller (5, p. 202), in order to subsume repression under their general theory of anxiety learning, speak of a "response of stopping thinking," reinforced by anxiety reduction.

Unfortunately, S-R theorists are somewhat vague in discussing the nature of the not-response. Sometimes Hull seems to identify it with the absence of activity, at other times with the cessation of activity, and on still other occasions he seems to assert that sI_R is I_R conditioned to the stimulus situation (all italics ours):

Consequently there would arise the somewhat paradoxical phenomenon of a negative habit; i.e., a *habit of not doing something* (10, p. 282).

Stimuli and stimulus traces closely associated with the cessation of a given activity, and in

the presence of I_R from that response, become conditioned to this *particular non-activity* . . . (11, p. 75).

The organic process most closely preceding the drive reduction would be the *cessation of the activity itself* (11, p. 75).

. . . this *cessation of activity* would become conditioned . . . (10, p. 282).

Stimuli closely associated with the acquisition and accumulation of inhibitory potential (I_R) become conditioned to it . . . (10, p. 282).

Miller and Dollard refer to a tendency to stop an activity:

. . . Thus muscle strain and fatigue are drives constantly motivating the subject to *stop the response* he is making; escape from muscle strain and fatigue are ever present to reward *stopping*. Extinction occurs unless the effects of the drive of fatigue and consequent reward for *stopping* are overridden by the effects of other stronger drives and rewards (24, p. 40).

From the above array of quotations, no clear indication emerges as to just what it is that gets conditioned to the stimulus in sI_R . However, these statements do suggest a relatively small number of alternatives. The not-response (that which gets conditioned to S in sI_R) is either: (a) the *absence* of a particular activity; (b) the inhibition—*interruption or cessation*—of a response already in progress; or (c) I_R conditioned to the stimulus as a *learned drive*. We shall now examine each of these alternatives.

1. *The not-response is the absence of a particular activity.* According to this alternative, the sheer absence of a response is that which by I_R reduction will be associated with the stimulus. This conception is completely untenable. For, in this sense of not-responding, the animal is performing innumerable and indistinguishable not-responses all the time. Simultaneously with not pressing the lever in a Skinner box, he is also *not* running a maze, *not* jumping to a

black card, and *not* playing three-dimensional chess. As a matter of fact, the same infinite set of not-responses is also performed when he *is* pressing the lever. Since all of these not-responses occur at the time of I_R reduction and of reward, they should all be conditioned to the stimulus. This is clearly absurd.

2. *The not-response is the inhibition—interruption or cessation—of a response already in progress.* This alternative seems to be the one most frequently implied by S-R theorists. Here it is asserted that before the not-response can be evoked, the response proper must at least have begun; that is, the animal always starts to press the lever before he stops doing so. But, in extinction he eventually fails to respond altogether, and does not even start to make the response, at least overtly. Of course, one might again suggest that when no overt response is started, there is at least an implicit one present, so that the conditions for the elicitation of the not-response as here conceived are met. Such a conception, however, raises another problem.

If implicit responses are to be utilized in S-R reinforcement theory, many phenomena suggest that these responses must be capable of being extinguished. As we have already seen, the latent-extinction results suggest such an interpretation. Furthermore, Hull's recent treatment of secondary reinforcement (11) deals with this in terms of fractional responses. Since secondary rewards can be extinguished, the r_g 's must be capable of extinction. Finally, we believe that Hull's theory of problem solving (9) can be shown to require the possibility of extinguishing implicit responses. Thus, within the context of S-R reinforcement theory, r_g 's must be extinguishable, and since

they are conceived as formally akin to overt responses, their extinction must follow the same laws as those proposed for the latter.

How could such extinction take place? There would seem to be a need for a not- r_0 to counteract the r_0 . But, according to the present alternative, such a not- r_0 can only be evoked after the r_0 has been initiated. Once again, the complete elimination of the (implicit) response must be explained, not its interruption. Since we are already at the level of implicit responses, an even more implicit response would have to be set in motion for the purpose. This is an utterly unpalatable concept.

The present alternative, then, seems to be unsatisfactory. In order to explain the total elimination of responses, it must resort to the postulation of implicit responses. When called upon to account for the extinction of implicit responses, it becomes yet more strained.

3. *The not-response is reactive inhibition conditioned to the stimulus.* Hull (10) sometimes writes of sI_R as I_R which has been conditioned to a stimulus. This implies that I_R and sI_R are of the same nature, except that in the first case the inhibitory force is produced as the direct result of effector action, while in the second case the identical force is elicited by a conditioned stimulus.

Such an interpretation is formally similar to the theory of fear behavior suggested by Miller (23). He assumes that fear is an internal response, reflexly connected with pain, which can be conditioned to an originally neutral stimulus under suitable conditions. Kimble (13) has criticized the interpretation of sI_R in such terms. He argues that I_R should be treated as an intervening variable, and that responses, rather than intervening

variables, become connected to stimuli.

Kimble's criticism may not be a crucial one. S-R theorists have rarely hesitated to endow their intervening variables with appropriate properties, and could perhaps treat I_R as if it were a response (or rather, a not-response), which can be conditioned to stimuli to generate sI_R . Even this formulation, however, raises some problems. If sI_R is thought of as a conditioned I_R -response, then sI_R and I_R must have identical response properties. From this point of view, the inhibitory processes involved in sI_R and I_R must be the same in all respects save the manner in which they are aroused.

Thus, the dissipation of inhibition following the withdrawal of a conditioned inhibitor must follow the same temporal course as the dissipation of I_R , the response generalization of sI_R and I_R must be equivalent, and so on. We do not know whether S-R reinforcement theorists would be prepared to accept these consequences of the present alternative.

There remains yet a fourth possibility, that of equating the not-response with an actual activity antagonistic to the to-be-extinguished activity, i.e., making the not-response a bona fide response. The resulting theory of extinction would be rather close to the interference theories of Guthrie (7) and Wendt (34), and at the very least would have to face many of the objections that have been leveled against these. In the absence of any evidence that Hull and his co-workers had this possibility in mind, it will not be considered here.

PARADOXICAL DERIVATIONS FROM THE THEORY

We have tried to show that the S-R reinforcement theory of extinction

encounters serious empirical problems, and contains some important conceptual difficulties. One might argue that, despite their shortcomings, the postulates of the theory permit us to deduce a great number of phenomena of extinction which actually occur. Unfortunately, however, they also necessitate certain other predictions which are clearly false.

1. *Predictions regarding the course of learning and extinction.* Hull and his co-workers believe that habit strength becomes asymptotic to a maximum value, and they usually assume that it does not decay with time. Furthermore, they assert that I_R and sI_R result as a necessary consequence of the *evocation* of the response, regardless of the presence or absence of positive reinforcement. Withholding reinforcement leads to extinction only indirectly; when no further increase in reaction potential occurs, the inhibitory action of I_R and sI_R grow unopposed.

From these assumptions it follows that the ordinary learning curve should not be monotonically increasing, but instead should rise to a maximum and then eventually return to the base line.⁵ For, as the habit is

repeatedly reinforced, sE_R approaches its asymptote. Once this asymptote is approximated, further reinforcements cannot add any further effective increments to the habit strength. Only I_R and sI_R can then be generated to any extent. (That sI_R is not yet at its asymptote is obvious: since extinction has not yet occurred, sI_R must be capable of further growth.) This means that from here on, further reinforcements can only lead to a *decrement* in performance, and will eventually cause the total elimination of the response. A pause between trials may at first lead to some recovery due to I_R dissipation, but this recovery will be short lived. Further trials must add to sI_R until it is approximately equal to sE_R , at which point no more recovery can take place. The learning curve will have reached the base line, never to come up again. Necessarily, then, there is no learned act which can be performed for any length of time; its very repetition—regardless of reinforcement—must lead to its eventual elimination.

This prediction is at odds with everything we know about the course of learning. The phenomenon of "inhibition of reinforcement" (8) occurs only under quite special conditions, and hardly begins to do justice to this deduction. The learning curve must return to *zero*, regardless of the spacing of trials, and must do so in the *same number of trials* required for experimental extinction after sE_R has reached asymptote. One does not have to refer to experimental studies to demonstrate the fallacy of this prediction. Our daily life is full of countless activities which we perform again and again with no sign of decrement. We turn door knobs, say "how do you do," sit down on chairs, and recline on beds, and have done so since childhood. It is reason-

⁵ Koch (15) notes this point in his review of Hull's *Principles of Behavior*, but does not seem to regard it as more than a matter of detail. The same problem is recognized by McGeoch and Irion (21, p. 55). They suggest that the situation could be remedied by making I subtract from N (the number of reinforced trials) rather than from sE_R . In effect, this would make I_R subtract from sH_R . In Hull's system, however, the indestructibility of sH_R and the merely "masking" roles of I_R and sI_R are essential; for example, they are crucial for the derivation of such phenomena as spontaneous recovery, reminiscence, and disinhibition. The suggestion made by McGeoch and Irion thus amounts to a proposal for a radically revised theory, and is not specifically relevant to the present discussion.

able to assume that such habits have reached asymptotic strength at an early age, yet there is no sign of decline.

By the same reasoning, it also follows that once a habit has been completely extinguished, reconditioning is impossible. For again, assuming that the habit strength was at asymptote prior to extinction, further reinforcement—no matter how frequent or how spaced—cannot add to it. This also is contrary to experimental fact (2) and to common observation.

The deductions just developed lead one to suspect that there is some serious flaw in the postulates which generated them. It seems to us that the problems principally derive from the assumption that there is no qualitative difference between the learning and the extinction situations, and that nonreinforcement affects performance merely by preventing the further growth of habit strength. According to the theory, an extinction trial is but a learning trial without reward—or rather with decreased reward—since I_R reduction still furnishes some reinforcement. Withdrawal of reward produces no real change in the situation. I_R and sI_R are generated during learning as well as during extinction. We shall now try to show that this conception leads to yet further paradoxes.

2. *Predictions regarding the impossibility of either learning or extinction.* In the theory of extinction originally proposed by Hull, conditioned inhibition is a habit established by reinforcement due to I_R reduction. Whenever an animal performs a response, a not-response inevitably follows it. Just how the not-response is conceived is irrelevant, so long as it inevitably occurs subsequent to the bona fide response. During the per-

formance of the not-response, I_R dissipates. This results in need reduction, and in turn reinforces the connection between the stimulus situation and the not-response. Thus sI_R is built up. The not-response opposes the response, eventually leading to extinction.

If we accept this mechanism, we are faced with an unpleasant dilemma:

a. *Extinction is impossible.* Before not-responding, the animal must necessarily have responded. If I_R reduction is reinforcing, it should reinforce the response as well as the not-response. If it did, and to the same degree, extinction could not take place.

In discussing this problem, Hull (10, p. 301) refers to the gradient of reinforcement. He points out that the not-response is temporally more contiguous with the decrease in "nocuous" stimulation and thus to reinforcement, than is the bona fide response. In consequence, the former should be stamped in more strongly than the latter, i.e., the increment in sI_R should exceed the increment in sH_R . In this manner extinction could take place. But this solution forces us onto the other horn of the dilemma.

b. *Learning is impossible.* As we have already seen, according to the theory, the not-response follows the response both during learning and during extinction. After pressing the lever in a Skinner box, the animal must necessarily stop pressing the lever (perform the not-response). This occurs before he reaches for the food pellet. But since the gradient of reinforcement—invoked before to make extinction possible—applies here equally, the not-response should be conditioned more strongly to the stimulus situation than should the response itself. In that case, the response can never become effectively

established, since the increment in sI_R must always be greater than the increment in sH_R . Any increase in the amount of reinforcement would benefit the not-response proportionately more than the response itself. Thus, learning is impossible.

Hull considers this problem also, and suggests a possible solution. He argues that the reinforcement in many experimental situations is secondary in nature, and that "this secondary reinforcement, e.g., the click of the magazine, occurs *during* the contraction and *before* the relaxation" (10, p. 302). Since reinforcement preceding a response is relatively ineffective, Hull concludes that the response would receive a greater benefit from the secondary reinforcement than would the not-response, even though the former benefits less from primary reinforcement due to I_R reduction. In this way, the not-response might receive less total reinforcement than the response proper, and the sH_R increment might outweigh the increase in sI_R .

This suggestion seems inadequate, for there is no reason to assume that the response is accompanied by more secondary reinforcement than is the not-response. The effectiveness of secondary reinforcers is generally believed to be a function of their temporal proximity to primary need reduction. The not-response is necessarily closer to primary reinforcement than is the bona fide response. Hence the secondary reinforcement accompanying it should be more, rather than less. The occurrence of a consistent click at the time of the response is merely an artifact of a particular experimental condition; surely the rat would learn even if the click were made contiguous with the not-response.

We are thus left with a strange

spectacle: a theory of extinction, derived from principles of learning, which must deny either the existence of learning or of extinction. The assumption of continuity between the learning and extinction situations—the failure to allow for any qualitative change brought about by withdrawal of reward—appears less and less tenable.

SUMMARY

Any theory of learning must deal with the phenomena of extinction as well as those of habit formation. S-R reinforcement theory, as presented by Hull, is one of the most influential of modern learning theories. It thus seemed appropriate to examine critically his treatment of extinction. We have tried to show that it faces a number of serious difficulties. In particular:

1. Recent experiments in the field of "latent extinction" suggest that the actual performance of a response may not be necessary for its extinction.
2. Neither reactive nor conditioned inhibition is clearly or adequately conceptualized. In particular, the "habit of not responding" has never received a satisfactory definition.
3. Certain paradoxical consequences can be derived from the theory: Not only should the learning curve inevitably decline to its starting point with continuous reinforcement, but, in fact, learning should be impossible altogether.
4. Many of these difficulties stem from Hull's assumption that withdrawal of reward introduces nothing essentially new to the situation.

REFERENCES

1. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1929, 4, 113-134.

2. BROGDEN, W. J., LIPMAN, E. A., & CULLER, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.*, 1938, **51**, 109-117.
3. BUGELSKI, B. R., COYER, R. A., & ROGERS, W. A. A criticism of pre-acquisition and pre-extinction of expectancies. *J. exp. Psychol.*, 1952, **44**, 27-30.
4. DEESE, J. The extinction of a discrimination without performance of the choice response. *J. comp. physiol. Psychol.*, 1951, **44**, 362-366.
5. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
6. GUTHRIE, E. R. *The psychology of learning*. New York: Harper, 1935.
7. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
8. HOVLAND, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. nat. Acad. Sci.*, 1936, **22**, 430-433.
9. HULL, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solving. *Psychol. Rev.*, 1935, **42**, 219-245.
10. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
11. HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.
12. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
13. KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500-510.
14. KIMBLE, G. A. Transfer of work inhibition in motor learning. *J. exp. Psychol.*, 1952, **43**, 391-392.
15. KOCH, S. Review of Hull's *Principles of behavior*. *Psychol. Bull.*, 1944, **41**, 269-286.
16. KRECHEVSKY, I. A study of the continuity of the problem solving process. *Psychol. Rev.*, 1938, **45**, 107-133.
17. LASHLEY, K. S. Studies in cerebral functioning in learning. V. The retention of motor habits after destruction of the so-called motor areas in primates. *Arch. Neurol. Psychiat.*, Chicago, 1924, **12**, 249-276.
18. LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72-87.
19. LIBERMAN, A. M. The effect of interpolated activity on spontaneous recovery from experimental extinction. *J. exp. Psychol.*, 1944, **34**, 282-301.
20. LIBERMAN, A. M. The effect of differential extinction upon spontaneous recovery. *J. exp. Psychol.*, 1948, **38**, 722-733.
21. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
22. MEEHL, P. E., & MACCORQUODALE, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372-396.
23. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89-101.
24. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
25. PAVLOV, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.
26. RAZRAN, G. H. S. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264-297.
27. RITCHIE, B. F., AESCHLIMAN, B., & PEIRCE, P. Studies in spatial learning: VIII. Place performance and the acquisition of place dispositions. *J. comp. physiol. Psychol.*, 1950, **43**, 73-85.
28. SALDANHA, E. L., & BITTERMAN, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, **64**, 37-53.
29. SEWARD, J. P., & LEVY, N. Sign learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39**, 660-668.
30. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1-40.
31. SPENCE, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 690-729.
32. THISTLETHWAITE, D. An experimental test of a reinforcement interpretation of latent learning. *J. comp. physiol. Psychol.*, 1951, **44**, 431-441.
33. TOLMAN, E. C., RITCHIE, B. F., & KALISH, D. Studies in spatial learning. I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, **36**, 13-23.
34. WENDT, G. R. An interpretation of inhibition of conditioned reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258-281.

(Received December 10, 1952)

PUNISHMENT: I. THE AVOIDANCE HYPOTHESIS

JAMES A. DINSMOOR

Indiana University

A possible reason for the seeming neglect of the topic of punishment in contemporary behavioral research and in most of our handbook and textbook presentations may be found in the present entanglement of theoretical treatments. So confused is the current picture that Stone, in a recent review of the literature, was led to remark that "The task of resolving apparently conflicting results . . . is an all but impossible one" (32, pp. 197-198). Actually, however, I feel that there is an available formulation which can handle the bulk of the data and which can incorporate it within a more general descriptive framework without requiring new explanatory principles. I also believe that this formulation can be shown to be consistent, at least, with those special and seemingly contradictory instances which appear to have been widely cited precisely because of the difficulties which they offer for any form of systematic treatment. I am speaking of the proposition that the main effects of punishment may be attributed to the establishment of certain avoiding reactions which prevent the completion of the original behavioral sequence.

The general suggestion that the effects of punishment may be due to some form of interfering reaction is by no means a new one. It appears as early as 1932, in some rather incidental comments by Thorndike. At that time, Thorndike presented a summary of several studies which seemed to indicate that punishment had little or no effect on the preceding behavior. However, he recognized the necessity of providing some kind of an "escape clause" to

deal with those special, as it seemed to him, and anomalous cases where the punishment of one response *did* facilitate the elimination of this response and the acquisition of a nonpunished alternative. To deal with such observations he offered the suggestion that this effect was due, not to a direct weakening of the punished response itself—as he had previously postulated in the law of effect—but to the strengthening of the alternative reaction. "The person or animal is led by the annoying after-effect to do something else to the situation" (33, p. 311). Or later, "The idea of making [the] response or the impulse to make it then tends to arouse a memory of the punishment and fear, repulsion, or shame. This is relieved by making no response to the situation . . . or by making a response that is or seems opposite to the original response" (34, p. 80).

Stemming from Thorndike's approach, we have such later developments as Guthrie's contiguity interpretation (e.g., 11), Estes' "anxiety" state (9), and various references to "heightened tension" (13, pp. 245-246) or to inferred drives of "fear" or "anxiety" (e.g., 8, 16) which are said to be reduced by making an opposing or conflicting response. In particular, several authors have at least mentioned an avoidance interpretation, the fullest treatments being those by Dollard and Miller (8, pp. 75-76), Mowrer (16, pp. 91, 118, 154, 210, 262 ff.; 17), Mowrer and Kluckhohn (18, pp. 80-81), and Skinner (31, esp. pp. 188-189); but even these are obviously rather brief. Furthermore, no attempt has yet been made to lay a detailed and comprehen-

sive statement of the hypothesis alongside the published findings from empirical studies of punishment to see how well such a statement fits the known facts.

In this paper I will merely outline the hypothesis itself. First, I will review some of the empirical studies of secondary aversive stimulation and avoidance training in order to see what principles are required for their interpretation. Next, I will compare the experimental operations used in avoidance training with those used in a study of punishment, in a free responding situation. Finally, I will try to show what should or must happen when we apply an aversive stimulus following successive instances of a given response.

AVERSIVE STIMULI

Since the concept of an aversive stimulus is fundamental to subsequent discussion, I will begin by offering a definition. I have selected the word *aversive* both for the frequency of its appearance in the experimental literature and for the strength of its behavioral connotations in everyday usage. (In Webster's International Dictionary [2nd Ed.], for example, *aversion* is first defined as "act of turning away" and *avert* is further defined as "to cause to turn away" or "to ward off, or prevent, the occurrence or effects of.") I will use the word in a strictly functional or behavioral sense, with no reference to its subjective properties or to any assumed drive which might be said to be aroused or reduced by the presentation or removal, respectively, of the stimulus. It will refer to a class of stimuli which are suitable for studies of "escape training" (13) or "aversion" (14). The critical observation is that *the reduction or elimination of the stimulus increases the frequency or probability of the preceding behavioral se-*

quence—that is, that it is reinforcing to the subject.

For the naive organism, this classification apparently includes such stimulating events as immersion in water and certain intensities of light, sound, temperature, and electric shock. This is not to say, however, that an aversive stimulus cannot be stripped of its original properties or that these cannot be overlaid by other properties acquired through special training or instruction. In practice, most of the empirical studies on avoidance and punishment have been based on the administration of shock to rats, although occasional reference will be necessary to other stimuli or other organisms.

HOW NEUTRAL STIMULI BECOME AVERSIVE

What happens when a neutral or ineffective stimulus is paired with one which is already aversive to the subject, such as shock? A relatively clear and simple answer to this question may be found in an experiment by Brown and Jacobs (2, Experiment II). The apparatus consisted of two adjoining compartments, each with a shock grid as a floor, which were separated by a two-inch barrier surmounted by a guillotine-type door. In the first stage of the experiment, the experimental animals (rats) were each given ten presentations of a pulsating light and tone paired with a pulsating shock. Each presentation consisted of nine seconds of light and tone, overlapping with a final six seconds of shock. No systematic means of escape was provided during this stage of the experiment.

The second step was to test the functional properties which had been acquired by the light and the tone as a result of their pairing with shock. Forty trials were given. On each trial the door was raised and the light and

tone were presented without further shock. When the animal passed over the hurdle from one compartment to the other the light and tone were turned off and the door was lowered behind him. The time required for the animal to respond was measured on successive trials.

A group of control animals, which had not been shocked, were found to run somewhat less promptly from trial to trial. But the experimental animals ran more and more quickly; their latencies showed a rather sharp drop for the first 16 or 20 trials. Later, however, a slight, but significant, rise appeared. The early decline in latency shows that the removal of the light and tone was a reinforcing operation which strengthened the response of running from one compartment to the other. The final rise in latency presumably reflects the gradual loss which occurs in the effectiveness of secondary aversive stimulation when it is no longer paired with the primary stimulus.

An attempt has been made by Barlow (1) to specify more exactly what is the necessary temporal relation between the primary and secondary stimuli, and a related study has been conducted in avoidance training by Mowrer and Suter (16, pp. 280 ff.). These studies both suggest that the critical relationship is between some phase of the secondary stimulus and, more specifically, the beginning or onset of the primary stimulus, such as shock. Similarly, the effects of presenting the secondary stimulus without the accompanying shock have been isolated and separately investigated in experiments by Schoenfeld and Antonitis (25) and by Page and Hall (23). These studies indicate that such a stimulus loses its aversive character when it is no longer paired with the primary stimulus.

On the basis of several replications, then, the main fact seems to be reliably

established: that a neutral stimulus which is presented just prior to or overlapping with the administration of a primary aversive stimulus, like shock, acquires an aversive property in its own right and becomes what we may call a conditioned or secondary aversive stimulus. When we try to make use of this stimulus as a reinforcing agent, however, a difficulty arises. The reinforcing operation—terminating the stimulus without shock—is incompatible with the establishing and maintaining operation, pairing the stimulus with shock. When it is terminated, therefore, without being paired with the primary stimulus, our secondary stimulus gradually loses its effectiveness. The temporary nature of the secondary aversive property might seem to limit the role which these stimuli can play in the maintenance of behavior over an extended period of time. The difficulty is readily resolved, however, if the pairing is restored whenever the stimulus is weakened and the animal fails to respond within an arbitrary time limit. This is the basic paradigm for what is known as avoidance training.

HOW AVOIDING REACTIONS ARE MAINTAINED

In studies like those we have just been considering, two separate and distinct operations have been employed in successive phases of the experimental procedure: (a) the secondary stimulus is paired with the primary stimulus, and (b) the termination of the secondary stimulus is used to reinforce a selected response. In a simple and relatively effective form of avoidance training, these two procedures are interspersed or interwoven. At the beginning of each trial a secondary stimulus or "warning signal" is presented by the experimenter. If the animal makes the required response the signal is terminated; but

when the animal fails to respond within an arbitrary time limit, the primary stimulus is applied.

As an example of this form of training, let us take "Group III-On-Run" from an experiment by Mowrer and Lamoreaux (16, pp. 126 ff.; 20). The warning signal was a change in the pattern of illumination produced by turning on two overhead lamps and turning off a single lamp beneath the grid. Five seconds of grace were allowed in which the rat could run to the opposite end of the alley or shuttle box. If he did this, the signal was terminated, or changed back, and no shock was applied; if he did not, the stimulus was followed by two seconds of shock. On the first day of training, the two animals in this particular subgroup made the response only twice apiece in 10 trials; but on the eleventh and twelfth days, both animals ran on all 10 trials. Thus, these animals learned to run to the opposite end of the apparatus when the only direct reinforcement for this act was provided by the change from a pattern of stimulation that was otherwise followed by shock. The results were similar for other subgroups. Although the procedure which is used in avoidance training is relatively complicated, the general effects of this procedure can be predicted from a study of the way in which neutral stimuli are made aversive or stripped of their aversive character.

If the conditioning of avoiding responses is based on the termination of secondary stimuli, and if the effectiveness of these secondary stimuli is based on their pairing with the shock, it follows that some limit must be set on the frequency with which the avoiding response will be made. When the animal makes the required response for several trials in succession, the pairing operation is interrupted and the effects of successive stimulus-terminations should

dwindle. That this is indeed the case is suggested by more detailed observations of the subjects' behavior under fairly comparable conditions reported by Sheffield (26). Here the warning signal was a two-second tone (apparently of fixed duration), with the shock coming in the last tenth of a second. Guinea pigs were used as subjects. If the animal turned a rotating cage or activity wheel by 1 in. before the shock was due, the shock was omitted. "With successive omissions of the shock," Sheffield reports, "the amplitude of the conditioned response tended to decrease and latency to increase, until the animal failed to turn the wheel the required inch in the required time, and another shock was received. As training continued, more and more successive conditioned responses occurred without requiring [shock], but extinction between [shocks] continued throughout the training" (26, p. 171). The amplitude was restored, of course, and the latency cut down once more after a trial on which the animal failed to respond and the shock was actually administered.

AVOIDANCE TRAINING WITHOUT A SIGNAL

Actually, no independent warning signal need be presented by the experimenter. This is demonstrated by some data recently reported by Sidman (27, 28). In this study, one-fifth of a second shocks were administered to the rat at regular intervals of time ("shock-shock interval"). The animal was permitted to avoid or delay the shock, however, by pressing a bar or lever at one end of the chamber. When he did so, the shock was postponed for another interval—sometimes the same, sometimes different—which was timed from the beginning of each successive response ("response-shock interval"). Some fifty

animals were successfully conditioned by this procedure.

In an extension of his preliminary work, Sidman conducted three of his original animals through an extended series of training sessions at a variety of response-shock and shock-shock intervals. Regular changes in the rate of responding were obtained from each of these animals as a function of the length of either interval. In general, the more frequently the animals were shocked, the more frequently they responded. The rate of responding rose with shorter and shorter shock-shock intervals and, up to a point, with shorter and shorter response-shock intervals. At relatively short response-shock intervals, however, a new phenomenon appeared: the original function gave way, and a "delay-of-punishment" gradient took over. That is, the rate began to decline with very short intervals between the response and the subsequent shock, and the response tended to disappear when the shock followed almost immediately.

AN INTERPRETATION OF AVOIDANCE TRAINING

In interpreting Sidman's data (and later, the operation of punishment) we are forced, in order to construct a general or inclusive description, to make an appeal to stimuli which are: (a) not clearly specified; (b) not readily observed or recorded; and (c) most important, not under the direct control of the experimenter. We might, by analogy, call the stimuli which may be presented or withheld at the will of the experimenter "independent" stimuli and those which are produced by the subject without the intervention of the experimenter "dependent" or "intervening" stimuli, all in accord with the application of these adjectives to the noun "variable." If the reader objects

to an appeal to such stimuli, there are two ways in which he may place them, in a sense, under the control of the experimenter. First, he may invade the organism by surgical or pharmacological techniques and presumably segregate or insulate the subject, in a manner of speaking, from the consequences of his own behavior. Or, as an alternative, he may add or subtract presumably equivalent stimuli to or from his experimental operations, making them likewise contingent upon the subject's response. Then we can see what effect these substitute dependent stimuli have on the over-all pattern of behavior. This procedure might well be used to substantiate Sidman's interpretation of his results.

Sidman's interpretation may be paraphrased as follows. Any form of behavior other than pressing the bar will eventually be followed by shock. The dependent stimuli that accompany such behavior thereby acquire an aversive character, through their pairing with the primary stimulus. But pressing the bar is never immediately followed by shock if a reasonably long response-shock interval is used, and the stimulation which accompanies this form of response does not become aversive. Hence, whenever a bar press follows some response that has previously been shocked, it will be reinforced by the change from an aversive to a nonaversive pattern of stimulation. At first only a few of the possible forms of response may have been shocked, while other responses have not been so paired. At this stage, the bar pressing is not always reinforced, and the rate of pressing fluctuates. As the training continues, however, more and more of the animal's behavioral repertoire is shocked, with the result that the avoiding response is fairly regularly reinforced. A ceiling is finally imposed by the very success of the avoiding response itself,

as in other studies, which reduces the over-all frequency of shock and thereby limits the frequency of pairing and the effectiveness of a given change in stimulation.

We now see that the warning signal, which is usually provided in a study of avoidance training, plays a relatively subtle role. There are three possible relationships which tend to be confounded and which are very difficult to segregate in a given study:

1. As in Sidman's procedure, the stimuli produced by *S* himself in making responses other than that prescribed by *E* are to some extent correlated with the shock and presumably provide some basis for the reinforcement of the avoiding reaction. This may be one of the sources for the responding that occurs in the absence of the signal, between successive trials (e.g., 19, 21). But these stimuli are in actuality paired with the shock only when they are accompanied by the signal. Thus, the "true" or most effective secondary stimuli are a set of stimulus combinations or compounds (24), each including the warning signal as one of its elements. The effectiveness of these compounds presumably depends on the exact temporal relation between the signal and the shock (e.g., 16, pp. 280 ff.; 37).

2. Furthermore, the effectiveness of the training procedure seems to depend also on whether both elements of these compounds are simultaneously terminated by the avoiding response; if the signal element is terminated before or after the response, the change in stimulation produced by the response is somewhat smaller and less discriminable, and the reinforcement is less effective (16, pp. 84 ff.; 19, 21). In avoidance training, as such, the relation between the signal termination and the *response* is necessarily confounded with the temporal relation between the signal and

the *shock*; but it may be separated in a study of secondary aversive stimulation, where the operations of stimulus pairing and reinforcement have been segregated.

3. Finally, the warning signal may "set the occasion" for the (maximal) reinforcement of the avoiding response. That is, it seems to act like a cue (8) or discriminative stimulus (14, 29, 30). Early in training the rate of responding may be about the same in the presence of the signal or in its absence (between "trials"); but as the training continues, more responses occur in the presence of the signal and fewer occur in its absence, *provided that an opportunity is given for the animal to make these nonreinforced responses* (3)—i.e., for *extinction of responses in the absence of the signal*.

Mowrer and Lamoreaux have discussed this problem in somewhat similar terms, and conclude that it is the necessity for forming a discrimination of some type between the presence and absence of the signal which slows down the acquisition of avoiding responses under the customary procedures (21).

A COMPARISON OF PROCEDURES

Now that we have seen how the subject learns to prevent the arrival of the shock in a study of avoidance training, we are in a position to apply these principles to the inhibition or suppression of the response which is observed in most studies of punishment. Actually, the two procedures are so similar that it is difficult to find a justification for any major distinction in theoretical treatment. As Mowrer says, such a distinction is "far from parsimonious, not to say an outright contradiction" (17, p. 421). What distinctions there are, of course, arise from the fact that in avoidance training the experimenter selects the response that shall be re-

quired to *avoid* the shock; whereas in an experiment on punishment he specifies and records the response that *produces* the shock.

This does, however, lead to certain consequences which may be worth inspecting. First, in avoidance training the class of responses which are *not* followed by shock is extremely narrow; it includes but one form—that specified by the experimenter as the avoiding response. But in a free responding situation the class of responses which *will* eventually be paired with the shock is extremely broad, including anything else the animal might do. The breadth of definition for these two classes of response is also reflected in the initial frequencies of behavior: before conditioning, the frequency of the avoiding response is likely to be relatively low, to constitute a small part, quantitatively as well as qualitatively, of the animal's activity; the combined frequency of other forms of behavior will be relatively high.

The situation is reversed in a study of punishment. Here, it is the class of responses which *are* followed by shock, for example, which is limited to a single behavioral sequence or chain. And in the usual experimental situation the initial frequency of this sequence is very low, so low, in fact, that it is ordinarily necessary to provide some form of reinforcement to boost its rate to a level where the inhibition may readily be observed. But the class of responses which are *not* followed by shock is relatively broad, including any form of response which conflicts with members of the punished sequence; and these responses are already quite plentiful at the beginning of training.

Second, the experimenter has chosen different criteria for the administration of the shock in the two cases, and this alters the detailed response-shock con-

tingencies both for the avoiding responses and for the punished responses. In avoidance proper, he delivers the shock at regular intervals whenever the animal *fails* to make the required response. He does not specify the exact relationship between other forms of response and the arrival of the punishment. A given alternative, therefore, need not immediately or invariably be accompanied by shock unless this is continuous, as in simple escape training. Thus, a certain amount of time will be required before each of these responses has effectively been paired with the punishment.

But in a study of punishment itself, the shock is directly contingent upon *making* a particular response. The pairing can be made immediate and invariable, unless the experimenter himself wills it otherwise. Special schedules, such as delayed or intermittent punishment, may readily be imposed.

Similarly, in a study of avoidance training the experimenter not only decides that a certain form of response shall lead to avoidance of the shock, he also determines *how long* the shock will be postponed following this response, if it is not repeated—as illustrated by Sidman's "response-shock interval" or by the customary interval between trials in a signal study.

Not so with punishment: Here, the relationship between the avoiding responses and the shock is less direct, for it depends on what these do to the original sequence of behavior. The consequent variation in the delay of punishment has a selective effect on various forms and durations of avoiding response (27, 28, 36, 37). The way is opened for a "shaping up" or differentiation of the original avoiding behavior.

Finally, the observations are different. In an avoidance study the experimenter has defined, by his criterion for admin-

istering or withholding the shock, the form of the avoiding response. This turns out to be the only response which can readily be recorded, since the alternative forms have not been defined and a specified alternative may not be inclusive. In a punishment study, on the other hand, it is the punished response which has been defined, and the avoiding responses cannot readily be recorded. In either experiment something might be gained by recording one response as a representative of the class of behavior which is not ordinarily observed, although the frequency of any single response is likely to be too low, unless experimentally reinforced, to provide a very sensitive index to the remainder of the animal's behavior.

DISCRIMINATION OF THE AVOIDING REACTIONS

There is one special problem in applying avoidance theory to the action of punishment which has not been explicitly discussed by previous writers. The behavior which is punished constitutes only a small fraction, qualitatively speaking, of the animal's total repertoire. In a laboratory study, to be sure, this behavior may have been strengthened to such an extent by direct experimental reinforcement that it constitutes a relatively large proportion, *quantitatively* speaking, of the animal's activity. In this case, elements of the punished sequence will intrude at frequent intervals between the avoiding responses. Most of the animal's behavior should remain "relevant" to the punished sequence. If it is the pressing of a bar which is punished, for example, he should spend the bulk of his time in the vicinity of this bar. He will not "get very far," to judge from analogous data (15), from the punished act. Since under these circumstances the animal is

almost always in danger of being punished, no special timing or discrimination of his avoiding responses may be required.

But in this respect the laboratory does not necessarily mirror life. Outside of the laboratory a given sequence may not have such a degree of strength. The animal may spend much of his time in activity which is essentially "irrelevant" to the punished response, i.e., which shows no major change in frequency following the institution of punishment. The punished responses intrude only occasionally among a variety of other forms of behavior. If we assume that there is some over-all limit to the frequency of the avoiding responses, a discrimination seems to be necessary. For if discrete avoiding responses were interspersed at random, regardless of what the animal might be doing, they would not conflict temporally with the appearance of the response which is punished and should have relatively little effect on its frequency.

In order, then, to inhibit or suppress the punished response, the avoiding responses must in a sense *anticipate or forestall it* by arising at just the moment when this response itself would otherwise appear. They must, we might say, be correlated with its expected occurrence. No "expectancy" construct, however, is required. The problem is much the same as the problem of accounting for the proper timing of avoiding responses to a signal, so that they may appear just prior to the primary stimulus. And the answer to this problem, too, is quite analogous.

CHAINING

Neither our own everyday behavior nor the activity of one of our subjects in the laboratory is made up, in atomistic fashion, of a random series of dis-

crete and unrelated acts. Experimentally reinforced behavior, in particular, flows along in fairly orderly and regular sequences or "chains" (14, 30), as may be established by the most casual observation. Most of our laboratory records, it is true, depend on timing or tallying a single response, such as pressing a bar, turning to the right, or entering a goal box. We do not and cannot record and quantify everything that the animal does. This should not lead us, however, to ignore, where relevant, the fact that the behavior which we are studying in the "modified Skinner box," the T maze, or the runway actually consists of a continuous flow of activity from which we have rather arbitrarily abstracted a single, readily recorded element.

Again we are forced to consider stimuli which are not directly under the control of the experimenter, for each of the actions in a behavioral sequence has some effect on the current stimulation. An action may enlarge, contract, add, subtract, or otherwise alter some set of visual stimuli as the animal turns his head or moves about; it may bring him into physical contact with some object in his environment, such as a bar, a pellet, a barrier, or a wall; it may produce apparatus noises or bring new odors; or, as a minimum, it will normally produce a certain amount of proprioceptive stimulation. Although these stimuli arise in the chain as a natural consequence of the animal's own behavior, without any special intervention by the experimenter, we can largely duplicate their relationships to a particular response or their own interrelationships by direct experimental manipulation. Work of this sort has been conducted largely under the headings of discrimination training and secondary reinforcement.

DISCRIMINATIVE AND REINFORCING STIMULI

A given chain is completed and reinforced only when the necessary members occur in the proper sequence or order. It will not do, for example, for the animal to go through the motions of pressing a bar when it is at the opposite end of the cage, or to chew before the pellet is in the mouth. The function of signalling, so to speak, when to make a given response, or of "setting the occasion for" this response, is performed by the stimulus elements in the chain. There is a three-term relationship here: discriminative stimulus—response \rightarrow reinforcement. It is only in the presence of the discriminative stimulus, as Skinner has called it (30, 31), or S^D , that the next response in the chain is appropriate and actually leads to a reinforcing state of affairs. This relationship is probably well known to most of my readers and need not be labored here. Empirical demonstrations are numerous. They show that when the reinforcement of a given response—e.g., pressing a bar—is made to depend on the prior presence of a certain stimulus, be it wholly arbitrary, the animal comes to make the response quite promptly (4, 29, 30), or with increased frequency (5), when the stimulus appears, but fails to respond with any great frequency when this stimulus is absent.

One of the stimulus functions, then, which is crucial to the formation of a chain, is the acquisition by this stimulus of discriminative properties. In addition it would appear that such stimuli also acquire reinforcing properties, along with their discriminative role, so that they also serve to maintain the strength of the response which produces them (e.g., 10, 30). Although the acquisition and loss of this property seem to be governed by the same factors

which govern the acquisition and loss of a discriminative property (4, 7, 22, 39), we customarily refer to these stimuli—while exercising their reinforcing function—as secondary reinforcers.

Let us now consider what happens when an aversive stimulus like shock is applied as a punishment following some particular member of the chain. Again we have a three-term relationship: discriminative stimulus—response → aversive stimulation. The punished response follows upon its appropriate stimulus; the punishment itself follows upon the response; thus, *through the mediation of the animal's own behavior, aversive stimulation is paired or correlated rather specifically with the discriminative stimulus for the punished response*. If, furthermore, the entire chain is run off fairly regularly and fairly swiftly, the aversive stimulation may also follow rather closely upon some of the stimuli which appear earlier in the sequence. And finally, it is closely associated with whatever stimulation may arise during the execution of the punished act itself (12, 16, p. 262; 17). This reduces to the same analysis if we break down what we had hitherto regarded as a single act into a more detailed sequence or chain in its own right.

These stimuli, then, play a role which is similar to that of the "warning signal" in the conventional study of avoidance training. Patterns of stimulation which include these elements are more closely and more frequently paired with the shock, and should be more effective as aversive compounds; responses which terminate these elements should be maximally reinforced; and by "setting the occasion" for maximal reinforcement, these stimulus elements should serve as cues or discriminative stimuli for the avoiding responses. (We do, in fact, find that arbitrary stimuli which indicate the punishment or nonpunishment

of a given response *do* affect its frequency [6].) In a sense, then, these are not only discriminative and reinforcing stimuli for members of the chain but *discriminative and reinforcing (i.e., by their termination) stimuli for a corresponding set of avoiding reactions*.

DIFFERENTIATION OF THE AVOIDING REACTIONS

Just as the animal must learn to make his avoiding responses at the *time* when the punished response is about to occur, to make them temporally incompatible, so he must also learn to make his responses of such a *form* that they are physically or topographically incompatible with the punished response (or with earlier members of the chain). It is obvious how this occurs. The pairing between the discriminative stimuli and the punishment is mediated, as I have said, by the animal's own behavior on continuing with the chain and making the punished response. The avoiding responses are reinforced precisely *because* they are incompatible with the original sequence; otherwise, they too would be followed by shock. While we cannot specify the exact form which these new responses will take, we can, from our knowledge of the basis of their reinforcement, make some tentative predictions.

First, the animal may halt, "freeze," or hold a pose. This probably involves a fine-grain vacillation between incipient movements toward completing the chain and opposing movements which serve to restore the original position (38). There may be some tendency for the animal to hold these positions for longer and longer durations (12, Experiment 3), as this further delays the punishment. But this development will be limited by the strength of the original chain (35, 38), and the mean duration of these holding responses will

presumably reflect most of the variables which influence the original rate of the punished response.

Again, the animal may make responses which are incompatible with the next member of the chain and serve as digressions from the sequence. Even a slight delay in the completion of the chain may to some extent be reinforced, and a certain amount of seemingly pointless "boondoggling" may be expected, like the dilatory behavior of a small child heading for bed. The animal may scratch himself, stand on his hind legs, "wash his face," or push the sawdust about. If these responses do nothing to cancel the previous member of the chain, however, they may well be followed by immediate completion of the original sequence.

A certain premium is therefore placed on those forms of response which "undo" or cancel out one of the members of the sequence by an opposing movement or a reversal of the progression. The animal may let the bar come up again; he may drop or let go of some object; he may turn his head away from the visual stimuli; or he may withdraw bodily from the locus of the punishment. These responses remove the most important elements of the aversive pattern, namely, the discriminative stimuli for the next response in the chain. Furthermore, they "set him back" in his progress, so that he is forced to repeat one or more members of the chain to get back to the point where he was before. Thus, some differentiation of the *form* of the avoiding responses would seem likely, on the basis of selective reinforcement—variations in the temporal interval between the response and the punishment (27, 28, 36, 37). Lengthy sequences of incompatible responding, such as wandering to the opposite end of the cage, might be strengthened to some extent if the original sequence is

weak; but these too are limited by the tendency to return to the chain. If the punished behavior is relatively strong, they may even be "crowded out" by the combined interference resulting from the original chain plus more localized avoiding responses. The over-all situation is reminiscent of the "equilibrium" studied by Miller, Brown, and Lipofsky (in 15), although their analysis is limited to a somewhat specialized situation.

SUMMARY

By punishing an animal for making a given response (that is, by applying aversive stimulation), we can reduce its frequency of occurrence. The purpose of this paper has been to show how we can fit this observation into a more general theoretical framework without adding new and independent principles to our system. Accordingly, I have tried to deduce the main effects of punishment from the principles already demonstrated in studies of secondary aversive stimulation and avoidance training. In general, my hypothesis has run as follows: The punished response is not an isolated incident, *in vacuo*, but a member of some sequence or chain of responses which is linked together by a series of discriminative, and thereby secondary reinforcing, stimuli. The stimuli which come immediately before the punished response are paired by the response itself with the ensuing punishment. By virtue of this pairing, they gain an aversive property in their own right. Any form of behavior which is incompatible with some member of the chain and delays the completion of the sequence will be reinforced, and thereby conditioned and maintained, by the corresponding elimination or transformation of these conditioned or secondary aversive stimuli. These responses are functionally equivalent to

the responses which are investigated in a formal study of avoidance conditioning. The fruitfulness of this hypothesis may therefore be tested by a detailed comparison of the functional relations observed in studies of punishment and studies of avoidance training.

REFERENCES

1. BARLOW, J. A. Secondary motivation through classical conditioning: one trial nonmotor learning in the white rat. *Amer. Psychologist*, 1952, 7, 273. (Abstract)
2. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, 39, 747-759.
3. COPPOCK, H., & MOWRER, O. H. Intertrial responses as "rehearsal": a study of "overt thinking" in animals. *Amer. J. Psychol.*, 1947, 60, 608-616.
4. DINSMOOR, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, 40, 458-472.
5. DINSMOOR, J. A. The effect of periodic reinforcement of bar-pressing in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1951, 44, 354-361.
6. DINSMOOR, J. A. A discrimination based on punishment. *Quart. J. exp. Psychol.*, 1952, 4, 27-45.
7. DINSMOOR, J. A. Resistance to extinction following periodic reinforcement in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1952, 45, 31-35.
8. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
9. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, No. 3 (Whole No. 263).
10. FERSTER, C. B. Sustained behavior under delayed reinforcement. *J. exp. Psychol.*, 1953, 45, 218-224.
11. GUTHRIE, E. R. *The psychology of learning*. (Rev. Ed.) New York: Harper, 1952.
12. HEFFERLINE, R. F. An experimental study of avoidance. *Genet. Psychol. Monogr.*, 1950, 42, 231-334.
13. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
14. KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
15. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Pp. 431-465.
16. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
17. MOWRER, O. H. Motivation. *Annu. Rev. Psychol.*, 1952, 3, 419-438.
18. MOWRER, O. H., & KLUCKHOHN, C. Dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Pp. 69-135.
19. MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, 54, No. 5 (Whole No. 247).
20. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, 39, 29-50.
21. MOWRER, O. H., & LAMOREAUX, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, 58, 196-212.
22. NOTTERMAN, J. M. The interrelationships among aperiodic reinforcement, discrimination learning, and secondary reinforcement. *J. exp. Psychol.*, 1951, 41, 161-169.
23. PAGE, H. A., & HALL, J. F. Experimental extinction as a function of the prevention of a response. *J. comp. physiol. Psychol.*, 1953, 46, 33-34.
24. SCHOENFELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. J. Hoch & J. Zubin (Eds.), *Anxiety*. New York: Grune & Stratton, 1950. Pp. 70-99.
25. SCHOENFELD, W. N., & ANTONITIS, J. J. A function of respondents in the extinction of operant responses. *Conf. exp. Anal. Behav.—Notes*, 1949, No. 17. (Mimeo.)
26. SHEFFIELD, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, 41, 165-177.
27. SIDMAN, M. Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science*, 1953, 118, 157-158.
28. SIDMAN, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *J. comp. physiol. Psychol.*, 1953, 46, 253-261.

29. SKINNER, B. F. The rate of establishment of a discrimination. *J. gen. Psychol.*, 1933, **9**, 302-350.
30. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
31. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
32. STONE, G. R. The effect of negative incentives in serial learning: II. Incentive intensity and response variability. *J. gen. Psychol.*, 1950, **42**, 179-224.
33. THORNDIKE, E. L. *The fundamentals of learning*. New York: Teachers Coll., 1932.
34. THORNDIKE, E. L. *The psychology of wants, interests, and attitudes*. New York: Appleton-Century, 1935.
35. TOLCOTT, M. A. Conflict: a study of some interactions between appetite and aversion in the white rat. *Genet. Psychol. Monogr.*, 1948, **38**, 83-142.
36. WARDEN, C. J., & DIAMOND, S. A preliminary study of the effect of delayed punishment on learning in the white rat. *J. genet. Psychol.*, 1931, **39**, 455-461.
37. WARNER, L. H. The association span of the white rat. *J. genet. Psychol.*, 1932, **41**, 57-90.
38. WINNICK, WILMA A. A study of incipient movements in avoidance. Unpublished doctor's dissertation, Columbia Univer., 1950.
39. WYCKOFF, L. B. The role of observing responses in discrimination learning. Unpublished doctor's dissertation, Indiana Univer., 1951.

(Received February 4, 1953)

THE MEASUREMENT OF VALUES¹

L. L. THURSTONE

University of North Carolina

In this paper I shall try to summarize briefly the attempts of several investigators to extend the concepts of measurement to the subjective domain. While this work is admittedly crude and exploratory, the results do look promising so that this field should be challenging for further study. Here we shall give only brief statements of the fundamental ideas without details of theory or experimental procedure. Our purpose here is only to sketch the nature of this field of research.

When we propose to measure human values, colleagues in the humanities may shudder at the very idea. When I wrote a paper entitled "Attitudes Can Be Measured," some of my colleagues did shudder. They were sure that social attitudes contain some essence that could not be identified and measured. They were sure that, in making the attempt, we would measure only the trivial.

Human values are essentially subjective. They can certainly not be adequately represented by physical objects. Their intensities or magnitudes cannot be represented by physical measurement. At the very start we are faced with the problem of establishing a subjective metric. This is the central theme in modern psychophysics in its many applications to the measurement of social values, moral values, and esthetic values. Exactly the same problem reappears in the measurement of utility in economics.

In order to establish a subjective metric we must have a subjective unit of

measurement. Before we can accept a subjective metric, it must satisfy the logical requirements of measurement as distinguished from rank order. These objectives have been approximated in the equation of comparative judgment and its variants.

Before proceeding to discuss the many applications of the subjective metric, we shall review briefly the principal psychophysical concepts by which a subjective metric can be established.

Let us consider these concepts in terms of a rather simple example, namely, the judgment of excellence of handwriting. When we look at several specimens of handwriting, it is fairly easy to select some that are considered to be excellent and others that are judged to be poor. In general, there is good agreement in such judgments. If we were asked to equate our judgments of excellence in a handwriting specimen to some physical measurements on the script, we would find it difficult. One of the main requirements of a truly subjective metric is that it shall be entirely independent of all physical measurement. In freeing ourselves completely from physical measurement, we are also free to experiment with esthetic objects and with many other types of stimuli to which there does not correspond any known physical measurement.

If we present a single handwriting specimen to a subject with the request that he tell us how good he thinks it is, then he must try to convey the degree of excellence in terms of words. It is well known that people vary tremendously in their use of superlatives in appraisals of experience, and, consequently, it is preferable to avoid such

¹ This paper was read at the Southern Society of Philosophy and Psychology in Knoxville, Tennessee, on April 11, 1952.

a direct procedure. Next we proceed to pairs of stimuli. We can ask the subject to judge which is the better of two specimens. In so doing, the subject gives his comparative judgment for each pair and he is not asked to give any verbal description of excellence.

The degree of excellence of a handwriting specimen is experienced by the subject in terms of some subjective process or *quale*. Since nothing is known about the neurological correlates of judgments of excellence of handwriting, we shall dodge all such terminology by merely referring to the discrimininal processes by which the subject does, in fact, discriminate between the different specimens. These processes may be assumed to be physical or truly subjective according to the preferences of the investigator. His preference on this point has nothing to do with the subsequent development of the law of comparative judgment.

When the subject makes a judgment that one specimen seems to him to be better than another specimen, we postulate discrimininal processes which differ in some manner in terms of which the percipient does make the discrimination. The more excellent specimen has some *quale* which differs from that of the poorer specimen. Imagine that the discrimininal processes which correspond to different values are arranged in a spectrum from those discrimininal processes in terms of which the percipient experiences the good specimens to the other end of the spectrum with discrimininal processes in terms of which he experiences what he calls the poorer specimens.

Consider next the phenomena of dispersion. If one subject were to examine the same specimen in comparative situations on a large number of occasions, it is not to be expected that he would always experience a particular specimen with the same discrimininal process. It

can be assumed that the same specimen will be experienced in terms of discrimininal processes in the same general region of the subjective continuum that has been postulated. So far we have no metric.

At this point we recall one of the fundamental restrictions on the problem of establishing a subjective metric. The discrimininal processes must be assumed to be of such a character that they do not necessarily have intensities or magnitudes which can be in any sense measured. This is an old problem that was discussed many years ago in psychophysical theory. For theoretical considerations, imagine that the discrimininal processes could actually be identified on each occasion when the subject makes a comparative judgment. The repeated observations of the same specimen can be assumed to produce an error variation from one occasion to the next. If we consider the relative frequencies of these discrimininal processes as responses to the same stimulus, then we can postulate a Gaussian error distribution for the responses to the repeated observations of the same stimulus. Let us now assume that the spectrum of discrimininal processes is stretched or contracted in different parts in such a way that the frequency distribution of these processes is Gaussian in terms of any given stimulus. Now we have a metric, but it is so far an entirely arbitrary metric. Imagine, at least in theory, that the same procedure can be repeated for many different stimuli which cover the whole range of discrimininal processes in terms of which degree of excellence is experienced. It is now a question of experimental fact whether the metrics determined for the separate stimuli will be the same when all of the stimuli are considered together. It has been found in many experiments that such is the case.

If we represent in the same model the

comparative judgment of two stimuli in which the subject says for each presentation which of the pair is the better, then we can observe the proportion of attempts in which the subject judges specimen j to be better than specimen k . If we have a whole table of such proportions, it is possible to infer the spatial separations of the different distributions of discriminial processes. Each stimulus is then assumed to project a Gaussian distribution on the subjective continuum with a mean and a discriminial dispersion. An ambiguous stimulus will project a wider dispersion on the subjective continuum than a sharply defined or relatively unambiguous stimulus. Each stimulus will then be defined in the subjective continuum by its mean position which is called a scale value and by the standard deviation of its dispersion of discriminial processes. Each stimulus is then defined by two parameters in the subjective continuum.

Before we can put numbers into these parameters, we must define an arbitrary origin which may be taken as the mean value that one of the stimuli projects on the continuum. As a unit of measurement we may choose arbitrarily the standard deviation of the dispersion which that stimulus projects on the subjective continuum. When that has been done, similar numerical values can be assigned to all of the other specimens that have entered into the comparative judgments. Further, we can test for the internal consistency of this theoretical model.

It should be carefully noted that we have not assumed that the discriminial processes have magnitudes of any kind. They have been dealt with merely as subjective quales and we have assumed only that in principle their relative frequency of association with any given stimulus can be ascertained. While this cannot be done directly, these frequencies can be inferred indirectly from the

observed comparative data. It should also be noted that we have not postulated the existence of any physical measures of any kind for the stimuli that have entered into the comparative judgments.

With this formulation of the law of comparative judgment, we are free to proceed with comparative studies of all kinds of stimuli which have no physical measure whatever. Hence we can turn to a wide array of interesting psychological problems involving value judgments. The freedom from any postulated physical measurement is the key that makes studies of this kind possible.

The method of comparative judgment turns out to be a rather general experimental procedure, and the well-known constant method in psychophysics is a special case in which one of the stimuli is arbitrarily taken as the standard which is compared with all of the other stimuli. Classical psychophysics was concerned with the more restricted problem of limen determinations.

We turn next to a brief review of some of the classical psychophysical methods because some of them have application in modern problems which transcend the determination of limens. In the method of equal-appearing intervals, the subject is asked to sort a large number of stimuli into a specified number of successive categories, say six or eight or ten. He is instructed to sort them in such a way that the intervals represented by the categories seem to him to be equal. This method is useful for rough survey purposes, but it can be shown that, even when the subject attempts to do this, he actually does not succeed in making the intervals subjectively equal. The method is, however, useful for coarse scaling such as the construction of attitude scales. The old method of equal-appearing intervals has been modified into what we call the method of successive intervals, in which

the intervals are defined by descriptive phrases or by sample specimens. This method has been found to be very useful in various types of surveys to be discussed.

One of the old psychophysical methods was to ask the subject to sort a number of specimens into rank order. It has been found that rank orders can be analyzed in such a way as to obtain data approximately equivalent to that of the method of paired comparison. The method of successive intervals can even be analyzed as a variant of the method of single stimuli.

Since Weber's law and Fechner's law have figured so prominently in the history of psychophysics, we shall make a few comments about these two laws in relation to the modern setting. These two laws are frequently referred to as the Weber-Fechner law with the implication that they are the same law, but that is an error. It is possible to set up experiments with rather simple stimuli in which one of these laws will be verified when the other one is not verified. It would be useful to set up such experiments in order to show clearly the separation between the two laws. Weber's law states that the proportion of judgments $R > kR$ is a constant. R signifies here the physical magnitude of the stimulus and k represents another constant. Weber's law is concerned solely with physical measurements. It does not explicitly refer to the subjective continuum. On the other hand, Fechner's law states frankly the relation between the subjective continuum and the physical stimulus continuum. Fechner's law states that this relation is generally logarithmic, and it should be taken as a rough approximation to the relation between the subjective and the physical continua. Further, it can be seen that Fechner's law is applicable only to those stimuli which have a physical magnitude as well as an ex-

perienced intensity. The law of comparative judgment is completely independent of any physical stimulus magnitudes. The problem of the stimulus error is not ordinarily of serious concern to our problem. It deals with the ambiguity in the mind of the subject when he is asked to judge a stimulus as to the intensity of the subjective experience. Sometimes he attempts instead to judge the physical magnitude. A good example is that of a grocery clerk who can judge the weight of a bag of sugar. If he were asked to serve as a subject in the method of mean gradation, he would probably commit what Titchener would have called the stimulus error. In the measurement of social values, we are not interested in physical measurements because in general they do not exist for such values.

A very important advance in the application of psychophysical methods was accomplished by Richardson when he devised the triad method for studying the dimensionality of a domain. Instead of asking a subject to judge whether one stimulus is x -er than some other stimulus where x is any specified attribute, he set up the discrimination experiment in such a way that no attribute was specified. In the method of triads, the subject would be shown three patches of color, for example, and he would be asked to indicate which is the odd one with the implication that the remaining pair is more alike than any other of the three pairs. In this way the subject can make judgments of the degree of similarity or difference without having any specified attribute. Data collected in this manner can be transformed into the equation of comparative judgment and the dimensionality of the domain can then be ascertained by the Young-Householder theorem. Such a method can be used experimentally to determine the dimen-

sionality of the various sensory modalities.

Perhaps the best known application of these experimental methods for the study of values is in the measurement of social attitudes. The most sensitive experimental procedure is to present the subject with pairs about which he is asked to make certain judgments. For example, he may be presented with pairs of nationalities, and he may be asked to judge for each pair which he would rather associate with. That type of experiment has been carried out in several ways. The judgments that are made by the subject depend, of course, partly on his own preferences which are closely related to his own nationality, and the judgments are also determined by the nationalities that are judged. If two groups of subjects are asked to make judgments of this kind, one can say on the basis of objective evidence which of the two groups is more tolerant of other nationalities. At one extreme we would have people who are completely tolerant toward all nationalities. They would then also, of course, be completely indifferent about their own. Such people would have no national loyalty or identification. At the other extreme we would have people who are said to be strongly prejudiced or biased. They would have extreme loyalties to some nationalities and extreme dislikes for others. I doubt whether we should consider either of these two extremes to be ideal.

Some years ago Fred Eggen wrote a master's thesis in psychology before he went into the field of anthropology. In that master's thesis he wanted to know the effect of different forms of question with reference to nationalities. He had five different questions representing different degrees of intimacy. All five groups of subjects were given the same lists of pairs of nationalities, but there were different questions. One group

had the question, Which of each pair of nationalities would you rather associate with? Another group had the same nationality lists, but they were given the question, Which would you rather have as a fellow student? Another group had the question, Which nationality would you rather have your sister marry? The proportions were superficially quite different, but the rank orders of the nationalities were essentially the same. In this case, we would probably find that the form of the question has a tremendous effect on the discriminial dispersion but relatively little effect on the order of the nationalities. The effectiveness of comparative judgment for studies of this type should be exploited further.

In studying the measurement of social attitudes, the attempt is sometimes made to validate such experiments in terms of overt behavior, but that is an error. Samuel A. Stouffer of Harvard wrote a doctor's dissertation some years ago at the University of Chicago on this problem. He investigated social attitudes by means of statement scales in reference to the prohibition issue. He obtained data about his subjects as to their actual behavior on prohibition. He found that there was pretty fair agreement between what the subjects said on the attitude scales and how they actually behaved. I should like to point out that, while such a comparison is of considerable interest, it is not a validation of the attitude scale. A man may be entirely consistent in what he says and in what he does about a controversial issue, and yet both of these indices may be dead wrong in reflecting his attitude. In order to determine a man's attitudes in the sense of affective disposition about a controversial issue, it will be necessary for his friends to ask him privately when he is free to speak his mind and when he is not likely to be quoted. His personal atti-

tudes may or may not agree with what he says and what he does. Here again, attitudes are essentially subjective experiences which may or may not conform with overt action.

Another distinction in the study of social attitudes which is sometimes lost sight of is that the cognitive and the affective appraisals may be entirely independent. For example, a group of subjects may agree in their strong dislike of communism. Someone might give them an examination in order to show that the subjects actually do not know what they are talking about. That might very well be true, but the psychological fact is nevertheless inescapable that the affective attitudes may be strongly for or against a stimulus even if there is a great deal of confusion about its cognitive description.

The statement scale is not so sensitive as the paired-comparison procedure. It consists in a set of statements to which the subject responds by acceptance or rejection of each statement. In constructing such a scale, one presents a large number of statements to a group of subjects whose principal qualification is that they can read English. These subjects are asked to indicate for pairs of statements which represents the stronger attitude for or against x , where x represents the psychological object to which the attitude scale refers. For rough survey purposes, the attitude scales are useful.

An interesting application of these methods of studying values is to appraise the effects of propaganda. We made a large number of experiments on the effects of motion-picture films on the social attitudes of high school children. Statement scales and paired-comparison schedules of various kinds were given before and after the showing of a motion picture. By this method we were able to ascertain whether a given picture had a significant effect and in

what direction it did affect the children's social attitudes.

The method has also been applied in the study of international tensions by noting newspaper editorials. In one of those investigations a study was made with Chinese and Japanese newspaper editorials concerning each other, and it was shown, by treating key statements from the newspaper editorials, that the tensions increased at a very great rate before the two countries were at war. Quincy Wright has suggested in his political science studies that such applications of psychophysical methods might be useful in studying international tensions before they become very marked.

An application of these subjective measurement methods which has not yet been made will be in the definition of the morale of a group. In general, the morale of a group is described by newspaper reporters and by others who mix their own value judgments with the characteristics of the group to be described. For scientific work we should have a definition of morale which is entirely independent of the value judgments of the observer. Such a definition could be stated in terms of the dispersions of all of the debatable issues within the group. Other applications would be in the comparison of cultural and nationality differences as to the values that are considered to be essential. It is unfortunate that most students of social psychology and political science are too descriptively minded to adapt the quantitative methods that may be available.

Let us turn next to the experimental study of moral values. We have carried out several experiments in which a group of subjects was given a list of offenses that were presented in pairs. For each pair the subjects were asked to indicate which of the pair they considered to be the more serious. On the basis of data of this kind and with the

aid of the equation of comparative judgment, we ascertain the scale values and dispersions for these offenses. In one case we gave a group of high school students such a list of offenses and we determined the scale values and dispersions for these stimuli for three occasions. The first presentation was a day or two before they saw a film that described the life of a gambler. A few days after seeing the film they were given the second similar schedule. About six months later they were given the third schedule. The film described the life of a gambler and we wanted to know whether this film had an appreciable effect on the attitudes of the high school youngsters toward gambling. We found that they considered gambling to be a much more serious offense after seeing this film than they did before seeing the film. In a number of experiments of this type, we also found that the motion pictures had much more lasting effects than is ordinarily supposed. In many cases we found that only half of the effect of the film wore off in six months. It should be said, however, that these experiments were carried out in small towns in Illinois where the children do not see so many movies as in the large cities. We carried out a similar experiment in the Hyde Park High School in Chicago where the children were given free tickets to a movie at the Tower Theater, a few blocks away. There we found that the effect was very slight. Our interpretation was that one movie more or less for children in a large city high school makes very little difference in their attitudes. These methods of studying moral values could be used very effectively in the comparison of different groups in a large city. The groups might represent different nationality backgrounds and different religious backgrounds. It would be interesting to ascertain what these differences would

be. Such social psychological studies would help us to understand the problems of the extremely heterogeneous populations in the large cities. In a similar manner we have investigated experimentally the summation effect in propaganda where the effect of a single stimulus does not show a statistically significant effect.

Another interesting field of application is in experimental semantics. It would be useful, for example, to have an index of affective intensity for adjectives in a dictionary. Two adjectives may be equivalent as to cognitive meaning and yet differ widely in affective meaning. The words famous and notorious might be examples. So are the words pleasant, gay, and hilarious. Such affective indices would be useful in translating a foreign language.

We turn now to another type of psychophysical problem. In the psychophysical methods that we have considered so far the main problem was to allocate each idea or object to a subjective continuum which may be unidimensional or multidimensional depending on the nature of the problem. In most problems it is unidimensional. For example, if we ask subjects to judge the relative seriousness of offenses, we are dealing frankly with a unidimensional continuum, even though the discriminations may take place in a multidimensional continuum. We have here an obverse psychophysical problem. Having determined the subjective space which describes a group of subjects as to their attitudes in some field, we now inquire whether we can predict in any way what these people will do. When we turn the psychophysical problem in this manner, we find some exceedingly interesting psychophysical theorems of a new kind. I shall give a few examples.

Consider two political candidates for an election. Let one of them have a

wide dispersion on the affective continuum. By this we mean that some people are very enthusiastic about this candidate, whereas others actually hate him. Let the other candidate have the same average popularity, but assume that he has a narrow dispersion so that very few people are enthusiastic about him and very few people strongly dislike him. If these two candidates come to an election, we should expect them to split the vote evenly. However, the more variable of these two candidates might introduce a third candidate of approximately equal popularity and who also has a narrow dispersion. Then we would have three candidates, one with wide dispersion on the affective continuum, and two candidates of narrow dispersion, and all three of them would be equally popular on the average. In such a situation, the more variable of the candidates would draw half the votes and the other two candidates would get twenty-five per cent each. These proportions would be altered somewhat depending on intercorrelations between the attitudes toward the candidates, but the principle can be illustrated in the general case for zero correlation. This principle is no doubt well known among politicians, but I doubt whether any of them have ever thought of this principle as a psychophysical theorem.

Let us turn to another simple example from the field of market research. Consider a mail-order house or a retail store which carries a limited number of neckties. They desire to please the majority of their clientele. The manufacturers offer many hundreds or thousands of necktie patterns. If you turn to market research people with this problem, they may ascertain the 20 or 30 or perhaps 50 of the most popular designs, and they may suggest that these be the designs that should be carried. But that is the wrong answer.

Suppose that several hundred necktie patterns were submitted to a sample of the clientele. With such records one could rather easily determine not only which patterns should be carried, but also the number of patterns that should be carried in order to satisfy a specified proportion of the clientele. We would start with the most popular design and set that aside to be included. In the sample population we would then eliminate all who chose that popular pattern. Then we would inquire about the most popular pattern in the remainder of the sample population. That pattern would be set aside as the second design to be accepted. Eliminating those who chose that pattern, we would ascertain the most popular pattern in the remainder of the sample population. Proceeding in this way, we would come to the point where an additional pattern would increase the selection by only a very small percentage of the population and that would be the time to stop. In such a procedure we could determine the number of patterns as well as the designs which should be used in order to satisfy a specified proportion of the clientele. The ordinary solution of selecting the most popular designs would lead to a situation where some customers are confused by having many patterns which are equally acceptable while other customers find nothing to please them. The maximum satisfaction will be derived by proceeding in some such way as I have outlined. There is nothing profound about this procedure, and yet it would probably be novel in market research. There are situations where problems of this sort can be of national importance. If it should be necessary to restrict the manufacture of civilian goods, then it might be important to encourage the manufacture of a limited number of designs for all sorts of things and to select those designs in such a manner as

to please the majority of the civilian population. In this manner the psychophysical methods may be important in contributing toward national morale.

Recently we made an experiment on the prediction of choice with regard to menus. In this problem we were concerned with the simplification of psychophysical methods to the point where they would be practicable for survey purposes. The psychophysical methods of the laboratory are often too laborious to be used in practical surveys. It was decided to adapt the method of successive intervals for this problem. We presented a list of 40 foods on a successive interval schedule in which each subject was asked to indicate by a single checkmark his relative degree of like or dislike for each food item. There were nine short descriptive phrases which represented degrees of like and dislike for foods. This schedule of 40 items required less than five minutes for each of several hundred adult men subjects. In addition to this short survey schedule, we also presented them with 16 menus in which they were asked to indicate what they would be likely to choose from each menu. For example, there were four lists of desserts, several lists of entrees, other lists of vegetables, and the like. For each menu the subjects were asked merely to check which they would select from a given list. Vanilla ice cream occurred in several of the dessert menus. The proportion of the subjects who select vanilla ice cream for dessert depends, of course, in part on their relative like or dislike for this dessert, but the selections would also depend on the competing items in the dessert list. By the application of the method of successive intervals and some theorems in psychophysics, we predicted the proportion of the subjects who would select each one of the items and there were 56 such predictions. These

predictions were based entirely on the short, five-minute schedule for the whole list of 40 foods. We compared these predictions with the actual choices that the subjects made when they were confronted with the actual menus. The agreement was remarkable. The maximum discrepancy was between 3 and 4 per cent with one conspicuous exception for a dichotomy, namely, roast beef and fried chicken. The ratings for these two items were both in the upper two categories and the discrepancy was there 8 per cent, which was probably due to the effect of coarse grouping. The experiment demonstrated quite adequately that the prediction of choice can be effectively made with very simple survey schedules if these schedules are properly analyzed.

Some of these experiments deal with rather trivial values while others deal with socially more important values, but our principal concern here is in the development of those scientific methods which can be adapted over a wide range of values whether they be socially important or trivial.

We turn next to the application of psychophysical theory to some experimental problems in economics. For a long time there has been considerable interest in the measurement of utility, but the measurements have generally been indirect. Psychologists have been able to measure utility experimentally for over two decades, but economists have not until very recently expressed interest in these methods. In the last few years there seems to have been a marked change in the attitude of economists to these problems. In principle, utilities can be measured for an individual subject, but it is easier experimentally to apply these methods to the measurement of utility for a group of subjects. Psychophysical theory lends itself well to a number of variations in the measurement of utility. For exam-

ple, the utility of a purchase can be described as the algebraic sum of the utilities of the object and of the price. In this case, the utility of the object would presumably be positive, whereas the utility of the price would be negative. The question then arises about the location of a rational zero point for the scale of utility. An experiment is now in progress to demonstrate an experimental procedure for locating the zero point in the scale of utility. It seems reasonable that the prices of various competing objects should be checked with their utilities to ascertain for any specified population to what extent some objects are overpriced or underpriced. Survey methods are available for doing these things. In determining the zero point for the scale of utility, we are asking several hundred subjects to express their preferences among various objects that might be given to them as birthday presents. Each of these single objects will then be given a value on the scale of utility. In addition to these judgments, we also asked the subjects to make a number of different judgments. We asked them whether they would prefer to receive gifts A and B or C. In this case they must judge whether the satisfaction from A and B is greater or less than the anticipated satisfaction from the single birthday present C. By judgments of this sort we expect to be able to locate the zero point of utility because the sum of the affective values of A and B combined should equal the utilities for these two objects taken separately. Within the range of the experiment with a small number of different objects to be selected, an additive theorem can be assumed to hold reasonably well. Diminishing returns would probably not be noticeable within the choice of four or five different objects.

In making these adaptations of psychological measurement theory to eco-

nomics, one naturally wonders whether economics could be developed as an experimental science. Although I am not an economist, it has seemed to me entirely feasible that economics should be developed as an experimental science. In discussing this question with some of my friends in economics, I find that they are divided. Some of them insist emphatically that economics can never be an experimental science, while others are equally certain that this is possible. As an example we might consider the indifference function in economic theory. An indifference curve can be considered as a curve showing the combinations of two commodities *X* and *Y* which have the same utility value. If the amounts of the two commodities are considered to be the *x* and *y* axes in a three-dimensional model, then utility can be considered as the ordinates which are perpendicular to the *x-y* plane. An indifference curve would then be a horizontal section parallel to the *x-y* plane which represents constant utility. For different values of utility we would then have sections at different elevations which give a family of indifference curves. It has been shown that these indifference curves can be determined experimentally. There are many situations of controlled economies where the shapes of these functions can be studied experimentally. Such situations are in occupied countries or in prisons and in other situations with central control of prices. By altering the price of a commodity, the changes in the indifference curves can be noted experimentally.

As a final example of the adaptation of psychophysical theory in the measurement of values, we shall consider the field of esthetics. If esthetics were to be regarded as a purely normative science, then we should expect the esthetic value of an object to be determined by its physical properties. Such an inter-

pretation seems well-nigh hopeless. It seems much more fruitful to recognize that the esthetic value of an object is determined entirely by what goes on in the mind of the percipient. In this manner of looking at the problem we deal again with values that are subjective experiences and which may vary from one person to another and certainly from one culture to another. An esthetic object symbolizes human emotional experience and its resolution in a conceptual and abstract manner. Except in extreme cases the esthetic experience is not itself emotional. It is essentially an abstraction. There is nothing absolute about the value of an esthetic object. The esthetic value is determined by the experience and the attitudes of the observer.

Some time ago I attended a series of seminars on esthetics at the home of one of my colleagues. Most of the participants in that seminar were from the humanities and the arts. The seminars were devoted to discussions about the theory of esthetics. In some of those discussions it occurred to me that the question at issue could be treated as a question of experimental fact, and I ventured to suggest how the psychophysical methods could be adapted to obtain an empirical answer to the question at issue. It was an illuminating experience to discover that some of my friends in the humanities were hostile to the very idea of subjecting questions of esthetic theory to empirical inquiry. On one of those occasions a friend showed me a quotation from Aristotle that settled the matter for him. It was heresy when I suggested that we knew more about this problem than Aristotle. Artists are sometimes suspicious of the experimental study of artistic preferences, and perhaps with some reason. Sometimes experimental studies are made in esthetics when the investigator is interested in secondary effects

rather than in the esthetic experience. On the other hand, I have found some artists who are very much interested in such inquiry. A friend who is a portrait painter frequently encouraged experimental studies of this kind at the Art Institute in Chicago. Unfortunately I have not been able to induce many students of psychology to study experimental esthetics.

In closing I should like to comment briefly on the social studies as science. It is unfortunate that the social studies have rather low prestige among the sciences. I believe that this is what we should expect because a large number of researchers in the social studies have not adopted the impartial, objective, and intellectual attitudes of science. Quite generally in these fields the writers argue for social action of some kind, about the right and wrong ways of life, about what is good and what is evil in the opinions of the writers, about the good and the bad names and categories for describing their political friends and enemies. It is still true that social scientists rather frequently fail to study social phenomena as science to identify the forces at work without name calling and without injecting their own value judgments into what they are describing. As long as social scientists fail to distinguish between propaganda and science they will have low prestige among the sciences.

SUMMARY

This paper has been concerned with the problems of a subjective metric. Social studies do not need to be quantitative in order to qualify as science. Some of the most important experiments in science deal first of all with the description of basic phenomena in a qualitative way. It usually happens that quantitative methods appear with more intensive study. Here we have con-

sidered some exploratory attempts to establish a subjective metric for the measurement of values. I have not succeeded in persuading social science students about the fascinating challenge to develop their field as science. To do so, we must free ourselves from the impulse for social action which has no place here. We should avoid problems in which we have an axe to grind. As

citizens we have the privilege and the duty to participate in political elections. But when we work as scientists we should be aloof from the issues of the moment and to the chatter of the market place. Only in scientific detachment and objectivity can we eventually be helpful in developing the social studies as science.

(Received April 8, 1953)

A NEURAL MODEL FOR SIGN-GESTALT THEORY¹

JAMES OLDS

Harvard University

Whether we like it or not, a theory of learning points two ways. In one direction it points to better experiments. In the other direction it points to a model that would reproduce the aspect of behavior which the theory is used to explain; it is the unfinished blueprint for such a model.

It is not so readily understood that the first pointing depends on the second; the theory must point to a model in order to point to better experiments. Quite often, because this is not understood, a further implication is overlooked, namely, the more nearly finished the blueprint, the better the experiments will be. I will try to justify this proposition briefly in the next paragraph, but first I would like to emphasize its main consequence for the present discussion. This is that "mechanical" or "neural" models are superior to merely "conceptual" ones because they do provide a more nearly finished blueprint. They tell us not only the type of relations that must occur, but the type of material in which these relations must occur, and how the relations can be built into this kind of material.

The advantage of the completely specified model or mechanism would be to allow synthetic reproduction of the phenomenon under investigation. Syn-

thetic reproduction gives the ideal solution to the main scientific problem: it apportions the variance of the phenomenon under investigation to the various causal constituents with no variance left over and not one too many causal constituents. Thus, it selects from the multitude of conditions that surround any phenomenon precisely the complex ingredients that are necessary to produce the phenomenon. In so doing, it gives the basis for a descriptive language that will not be crowded with irrelevant concepts, nor lacking in crucial ones, but rather will have just one concept for each important variable and none left over.

This would be the advantage of a completely specified model; the nearer we approach the completely specified model, the more we approach these advantages. Thus, it is to our advantage to get more specifications into the unfinished blueprint for the model. I believe the further implication is that an approach toward a mechanical model will always be beneficial.

THE ADEQUACY OF THE MODEL TO THE DATA

A model may fail, however, in either of two directions. On the one hand, it may be so incompletely specified as to fail to provide an adequate descriptive language and to carve out crucial variables. On the other hand, it may be more or less completely specified, but fail to reproduce the phenomenon under investigation.

My contention is that Hull's model (2, 3) is more completely specified than Tolman's (6); in this sense Hull has the edge. Tolman, on the other hand,

¹ This paper is based on portions of a dissertation submitted to Harvard University in partial fulfillment of the requirements for the Ph.D. degree in social psychology. The work was supported by a Research Training Fellowship of the Social Science Research Council and by funds of the Laboratory of Social Relations at Harvard. The writer wishes to express his appreciation to Professor R. L. Solomon for his many helpful criticisms and suggestions.

presents a model that seems to reproduce more adequately the phenomena of learning and performance that are the subject matter of both theories; in this sense Tolman has the edge. I want to consolidate their gains.

My purpose in the present paper, therefore, is to set forth a more complete blueprint for the model which Tolman has presented. I will do this by giving a neural interpretation of Tolman's theory based in large part on Hebb's (1) discussion of the properties of cell assemblies.

ADVANTAGES OF THE MODEL

As the proof of the pudding must be in the eating and not in any complicated rationalization, I will suggest at the end of this paper some of the advantages produced by the additions which I make to the Tolman theory. These come under three headings: (a) resolution of the problem of latent learning, (b) the stimulus control of ideas, and (c) the growth of approach motives. As it would do no good to expand on advantages before we have the theory, we proceed immediately to

an introduction of the various important points of the model.

Hebb's Cell Assembly

The cell assembly described by Hebb (1) is most simply conceived as a three-dimensional lattice of neural paths providing several complete circuits, and alternative paths from each junction point so that when an impulse finds one of the transmission units refractory, another path allows the impulse to stay alive within the system. Therefore, the system has the capacity to reverberate. The assembly is most easily understood on the basis of the diagram in Fig. 1 borrowed from Hebb (1, p. 73). Each of the arrows in the diagram represents a single transmission unit, a single pathway. Although these are not considered by Hebb to be individual neurones, but rather low-order systems of neurones, we will take them to be the lowest order of functional units for our present explanation. Each pathway is refractory for a moment after an impulse has traversed it. Therefore, without alternative pathways reverberation would quickly die out, for the impulse would come back a second time before a pathway could recover. Each cell assembly consists in a number of these paths; the diagram represents a cell assembly. From the diagram, we can see how alternative pathways make reverberation possible. The impulse enters along the pathway marked 1,4, it proceeds to 2,14, and then through 3,11 and 1,4 again. At this point, it finds 2,14 refractory, but there is an alternative path, 5,9. The impulse proceeds around according to the numbers and is allowed to stay alive within the system because *neither all the pathways, nor too many of them are refractory at the same time*.

Hebb's cell assembly as it stands has five properties that we should note before we proceed. The first is rever-

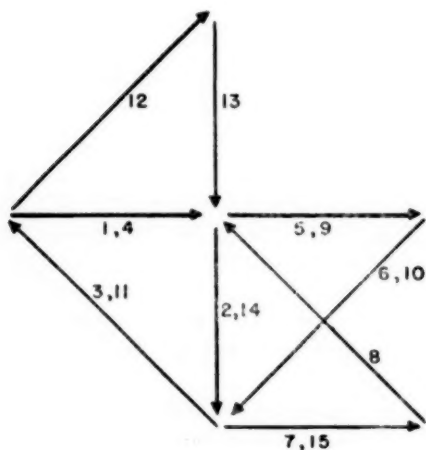


FIG. 1. The cell assembly described by Hebb (1, p. 73)

beration. When an impulse enters the assembly it can reverberate without further stimulation. Second, the cell assembly has relations to other internal assemblies so that it can be aroused by central facilitation. Third, it has relations to the peripheral receptors so that it can be aroused by the environment. Fourth, it tends to have behavioral outlets, that is, it tends to control behaviors while it is aroused. Fifth, it has at least two states or phases: it can be latent when it is not aroused, and it can be in a state of reverberation when it is aroused.

THE FOUR PHASES OF IDEAS AND WANTS

At this point we turn to the aspects of behavior that are to be explained by Hebb's construct. There are two enigmatic terms avoided by S-R psychologists and often by cognitive psychologists because they seem so subjective and unfathomable, and so particularly refractory to mechanical analysis. These are "ideas" and "wants."² No psychology is lacking a set of euphemisms for these terms, but few psychologies handle the problems well. S-R psychology speaks of "fractional components" instead of "ideas," and of "antedating goal reactions" instead of "wants." Tolman faces the problem with less circumlocution: he speaks of the "expectancy" or the "significate" instead of speaking of the "idea." And he speaks of the "readiness" or the "demand" instead of the "want."

If, instead of searching for better and more satisfactory euphemisms, we take the terms as they stand with their more or less obvious, everyday meanings, and ask what we know about them, we find

that we know quite a lot more than we might expect. And we also find that there is an interesting parallelism between an analysis of ideas and an analysis of wants that suggests that they are not such different things as they might seem at first glance.

The present analysis is going to be quite cursory and gross, for it is only to prepare the way for the model which is to come; it is to give some meaningful anchorage points for the technical material that is to follow.

First, there are various phases found in the analysis of a single idea. Let us take as an example the idea of a red light (of the traffic control variety). At first the red light is seven or eight blocks up the road, and we are not even thinking about it. I will say that the "idea of the red light" is *latent* at this point. After a few moments, we are approaching the intersection, the light at the corner turns from green to yellow, and for a very little while we are thinking about the red light and expecting it, but we are not seeing it. I will say the idea of the red light is now in a state of *expectancy*. But then the light turns red and we are seeing a red light. I will say the idea of the red light is in a state of *perception*. After we have sat behind the light for what seems like an interminable number of seconds, we become fed up with it, it seems to be lasting forever. I will say the idea is in a state of boredom. Finally, the light turns green, we drive on and forget it. The idea of the red light is latent again. It is obvious that an idea has at least four distinct conditions or phases: (a) it is not even thought, (b) it is thought but not seen, (c) it is seen, and (d) it is palling. The second condition can be divided again and again; when the idea is thought but not seen, it can be a mere thought, an expectancy, a memory, and so forth, but we will ignore these finer

² By the term "want" at this point, I refer to more than the basic physiological drives that underlie some (but not all) of behavior. Instead, I refer to the specific conceptualization of a goal that seems to precede most goal-directed activity in a human being.

gradations in the present paper. For our purposes, the idea has four phases which we may call latency, thought, perception, and boredom.

The most interesting thing about these four phases is that they are exactly and obviously paralleled by the four phases of a want. At first the want is latent; as for example when I am not thinking about food, and I do not want it. Next, something makes me think of food, and I notice that I am hungry. I start doing things that will get me fed; the want is now in a state of motivation. After that, I am being fed. The want is in a state of gratification. Finally, I am too full, and the want is in a state of satiation. After I have waited for a while, the satiation disappears, and the want is latent again. Thus, the want has four phases which we may call latency, motivation, gratification, and satiation. Note how closely these fit the phases of the idea.

From the parallelism, one would be tempted to suggest that ideas and wants are much the same sort of things. I suggest that they are not distinguished as far as the kind of structure is concerned, but only in terms of some power or "motive force" parameter. That is, an idea is a concept with a low motive force; a want is a concept with a high motive force.

THE FOUR PHASES OF THE CELL ASSEMBLY

The cell assembly, as we left it a few paragraphs back, has only two phases, latency and arousal. The state of arousal is a state of reverberation; an impulse enters the system along one pathway and reverberates within the system without further stimulus support.

We would like to find some characteristic of the cell assembly, implicit in

Hebb's description of it, to allow us to ascribe it four phases and thus use it as an adequate model for the ideas and wants we have just described. Particularly, we would like to find two different conditions of arousal, one corresponding to perception or gratification and the other corresponding to thought or motivation. Analyzing these two phases, we find that the thought-motivation phase is characterized by a minimum of external stimulus support: it is a more or less autonomous internal reverberation, and it does not seem to be terminated either of its own accord or by mere withdrawal of the arousing stimulus. Rather, this phase of expectancy or motivation is terminated by the presentation of the goal object in the environment.

The perception-gratification phase, on the other hand, is characterized by a maximum of external stimulus support: it is not an autonomous internal reverberation, it does seem to become satiated or refractory of its own accord, and it seems to go out immediately upon withdrawal of the arousing stimulus. This phase is the perception or enjoyment that is turned on by the goal object in the environment.

Our problem is this: How can the same idea participate in an expectancy which is terminated by the goal object, and in a perception which is turned on by the goal object? The same idea seems to be turned off and on by the same object, which sounds ridiculous.

I find the answer to this question in Hebb's discussion of the conditions necessary for reverberation, an answer which shows that Hebb's cell assembly is a much better model for ideas and wants than one might expect from a superficial glance.

You will remember that in our description of the cell assembly, we said it would reverberate because *neither all the pathways, nor too many of them*

are refractory at the same time. At the present point, this assertion becomes crucial. We may suggest that any stimulus which has a single or small number of connections with a given cell assembly would start a reverberation (a thought or motivation process in the assembly). A stimulus which has a large number of connections to many of the different pathways, on the other hand, would not set up a reverberation; instead it would "fire" the assembly. All pathways would be rendered refractory (or relatively refractory) at the same time. In the continued presence of the strong external stimulation the activity of the assembly could be maintained. But upon withdrawal of the external stimulus the assembly would be refractory, and activity would cease.

For our purposes, then, the cell assembly has four phases or conditions. We will say it can be in a state of latency, in a state of reverberation, in a state of firing, and in a state of refractoriness. These correspond to the four phases of ideas and wants. For the first phase we have used in all cases the term *latency*. For the second phase, we render equivalent the terms *thought*, *motivation*, and *reverberation*. For the third phase the equivalencies are *perception*, *gratification*, and *firing*. For the fourth phase the terms are *boredom*, *satiation*, and *refractoriness*. The cell assembly is our mechanical model for ideas and wants. A cell assembly of low "motive force" is an idea; a cell assembly of high "motive force" is a want. We will go on now to a simplified discussion of association.

THE ASSOCIATION OF IDEAS

Again we turn to the aspect of behavior that is to be explained, and again we find a phenomenon which is rarely treated in contemporary psychology except with careful circumlocution. This

is the association of ideas which is produced within a human being by a succession of stimuli in the environment. Each of us knows from his own experience a great deal about the way an associational link between two ideas functions, but we do not often analyze the functioning carefully enough to be aware of its essential characteristics.

I will take a simple example to make these characteristics explicit. Our subject is unacquainted with his typewriter. The carriage is far to the right, and he perceives and pushes a key marked "Tabular." The carriage jumps five spaces to the left and stops in a new position. First, there is an antecedent situation; then he makes a response and an outcome ensues. The antecedent situation is the carriage far to the right plus the perception of the tabular key; we will call this *A*. The response is to push the tabular key; we will call this *R*₁. The outcome is the carriage five spaces to the left; we will call this *B*. Thus, in the presence of *A*, *R*₁ leads to *B*. The *A-R*₁-*B* learning sequence has taught our subject that *A* followed by *R*₁ leads to *B*. We may say there is now an association of the *A* idea through *R*₁ to the *B* idea. In the future, if *B* is wanted, and *A* is presented, our subject will perform *R*₁. Also, if *A* is presented and *R*₁ should occur by accident, our subject will expect, and prepare for *B*. That is, if he wants the carriage moved from its *A* to its *B* position, he will now press the tabular key. And if he inadvertently presses the tabular key, he will expect and quite likely take some action to offset the movement of the carriage to its *B* position.

After a certain response in the presence of *A* has led to *B*, we say that some idea of *A* is associated with some idea of *B*. But the facts of behavior are these: (a) in the future if we make this particular response to *A* we will anticipate or expect *B*. (b) In the future if

we should happen to want *B* we would show some tendency to search out *A* and then to make this particular response that takes us from *A* to *B*. The perception of *A* now arouses some expectancy of *B*, and the motivation of *B* induces motivation of *A*. The link seems to carry expectancy in the *A* to *B* direction, and motivation in the *B* to *A* direction. This will become clearer now as we lay out the specifications for our model in detail.

A NEURAL MODEL FOR SIGN-GESTALT THEORY

There are two undefined structural units of the model. These are the *cell assembly* and the *response control unit*. We presume at the outset that for any stimulus with which the subject has repeated commerce, a cell assembly becomes established; thereafter, the stimulus is an *unconditioned stimulus* of the cell assembly. Further, we presume that for any response which becomes organized within the behavior repertory of a subject, a response control unit becomes formed; thereafter, the response is elicited by the activation of the response control unit. These two formative processes may occur at first more or less by chance; the rules of organization and growth given below will show how selectivity can be introduced after a chance generation of these structural units.

In the exposition, cell assemblies will be designated by the lower-case letters of the early part of the alphabet, e.g., *a*, *b*, *c*. Response control units will be designated *r*₁, *r*₂, *r*₃, and so forth. Stimuli in the environment will be designated by the upper-case letters of the early part of the alphabet, e.g., *A*, *B*, *C*. Responses will be designated *R*₁, *R*₂, *R*₃, and so forth.

The definitions or specifications and postulates are listed below.

I. The unconditioned stimulus. Each cell assembly has a *stimulus threshold of firing* which needs to be crossed by stimulation from the environment. A stimulus which crosses this threshold is an *unconditioned stimulus* of the assembly. The unconditioned stimulus of assembly *a* is *A*, that of *b* is *B*, and so forth.

II. The conditioned stimulus. Each cell assembly has a *stimulus threshold of reverberation* which needs to be crossed by stimulation from the environment (mediated by antecedent assemblies as noted in VII below). A stimulus which crosses this threshold is called a *conditioned stimulus* of the assembly.

III. The motive threshold. Each cell assembly has a *motive threshold* which must be crossed by *combined positive motive force* or *combined negative motive force* (see VIIIa and b below).

IV. Intrinsic motive force. Each cell assembly has an *intrinsic positive motive force* and an *intrinsic negative motive force* which contribute toward combined positive and negative motive force respectively (and toward the combined motive forces of its antecedents when it is reverberating, see VIII below). Thus, there are two separate force parameters of each cell assembly; it is as though there were a solution with two separately variable factors dissolved.

V. The law of assembly activation. Both the motive threshold and one of the stimulus thresholds must be crossed at the same time for the assembly to become aroused (i.e., to fire or reverberate). If the motive threshold is crossed, then:

(a) the assembly will *fire* if the stimulus threshold of firing is crossed. Arousal ceases upon termination of this stimulus.

(b) the assembly will *reverberate* if the stimulus threshold of reverberation is crossed (unless both stimulus thresholds are crossed, in which case the as-

sembly will fire). Reverberation continues after withdrawal of this stimulus; it is terminated by firing.

VI. The learning law of association. Two cell assemblies become related to one another by an associational relation under the following circumstances. If *a* fires, and then r_1 is activated, and then *b* fires, an associational relation will be formed between *a* and *b* which passes through the response control unit r_1 . The cell assembly *a* will become the antecedent of the associational relation, and the cell assembly *b* will become the successor of the associational relation. They will be connected with one another through r_1 . It is as though there were a wire connecting two terminal boxes *a* and *b* passing through a junction box r_1 ; and certain characteristics of the flow across the wire determine what will happen in the junction box (see IX below). The associational relation will be strengthened by further firings of *a* followed by activation of r_1 and firing of *b*. It will be weakened by further firings of *a* followed by activation of r_1 when these are not followed by firings of *b*.

VII. The law of conditioned stimuli. In the future, the firing of the antecedent will be a *conditioned stimulus* for the successor (see II and Vb above).

VIII. The law of the backflow of motive force. In the future, the reverberating of the successor will add two components of motive force, *instrumental positive motive force* and *instrumental negative motive force*, to the antecedent; these contribute toward respective combined positive and negative motive forces of the antecedent. A reverberating successor adds these components not only to the antecedent, but through the antecedent to further antecedents; the intervening assemblies need not be aroused for this transmission to continue to further antecedents.

(a) The *combined positive motive*

force of an assembly is equal to the sum of its *intrinsic positive motive force* and its *instrumental positive motive force*. Similarly, *combined negative motive force* is equal to the sum of *intrinsic and instrumental negative motive force*. Either the combined positive motive force or the combined negative motive force of an assembly must be above the motive threshold in order for the assembly to become activated (see III above).

(b) The instrumental motive force (positive or negative) which a reverberating successor delivers to a near or distant antecedent is: (i) directly proportional to the combined motive force of the successor, (ii) directly proportional to the strength of the weakest link in the chain of associational relations between them, and (iii) inversely proportional to the number of assemblies interpolated between them.

IX. The law of performance. The likelihood of a response R_1 depends on the amount of *facilitation* and the amount of *inhibition* contributed to the response control unit r_1 . Facilitation and inhibition are contributed to a response unit r_1 only when its antecedent *a* is firing and its successor *b* is reverberating. If *a* is firing and *b* is reverberating, then:

(a) Facilitation will be contributed to r_1 in proportion to the *amount of the difference* between the *combined motive force of the antecedent* and the *combined motive force of the successor* if this difference is *favorable to the successor*. Therefore, (i) if the successor is less negative than the antecedent, the response will be facilitated; (ii) if the successor is more positive than the antecedent, the response will be facilitated.

(b) Inhibition will be contributed to r_1 in proportion to the amount of the difference between the combined motive force of the antecedent and the combined motive force of the successor if

this difference is *favorable to the antecedent*. Therefore, (i) if the successor is more negative than the antecedent, the response will be actively inhibited; (ii) if the successor is less positive than the antecedent, the response will be actively inhibited.

(c) If the facilitation is greater than the inhibition, then the response control unit r_1 will be activated, and R_1 will occur. If the inhibition is greater than the facilitation, then r_1 will not be activated, and R_1 will not occur.

X. The law of motive growth and decline. The intrinsic positive or negative motive force of an assembly grows and declines as a function of variables. I suggest the following postulates as a program for research.

(a) The intrinsic positive or negative motive force of an assembly is a joint, direct function of the number of transmission units in the assembly (see *b* below) and the amount of positive or negative motive force internal to each transmission unit (see *c*, *d*, *e* below). Each transmission unit has both positive and negative motive force internal to it.

(b) The number of transmission units in an assembly tends to *increase* in proportion to the amount of time the assembly spends in a state of firing.

(c) The amount of positive or negative motive force internal to each transmission unit in the assembly tends to *decrease* in proportion to the amount of time that the assembly spends in firing.

(d) The amount of positive or negative motive force internal to each transmission unit in the assembly tends to *increase* in proportion to the amount of time that the assembly spends in reverberation.

(e) The rate of positive or negative motive growth during reverberation (see *d* above) will *increase* as a function of the *combined* positive or negative mo-

tive force of the assembly during the period of reverberation.

(f) The rate of positive or negative motive decline during firing (see *c* above) will *decrease* as a function of the *intrinsic* positive or negative motive force of the assembly during the period of firing.

INTERPRETATION OF TOLMAN'S THEORY

We turn now to sign-gestalt theory to show that our mechanical model does give interpretation to all of its important points. We will first interpret the chief terms of Tolman's theory; then we will show how the relations postulated by Tolman are inferences from our model.

Perception: this is a term which is not accented as basic by Tolman; implicitly, however, it has a very basic place in his theory. For it is not the presence of a stimulus in the environment which controls behavior, in the Tolman formulation, but the "perception" of the stimulus by the subject. Perception is always selective; stimuli are perceived in proportion to their relevance to motives (6, p. 35). Tolman defines a perception as "an expectation of the component of a sign gestalt when this expectation results primarily from *present* stimuli coming then and there" (6, p. 452). I believe this may be paraphrased simply by saying a perception is the apprehension of an object by a subject when this apprehension depends on immediate stimulation. Our mechanical analogy for perception is the firing of a cell assembly. It requires both the presentation of the unconditioned stimulus (V_a) and adequate combined motive force (V). The latter postulate accounts for the selectivity of perception.

Demand: this term is defined by Tolman as an "innate or acquired urge" to get to or from some given stimulus, or

some physiological quiescence or disturbance (6, p. 441). Simply, this is a want; it is an appetite or an aversion. A demand in our mechanical system consists in either one of two states. In the appetite case, it consists in the reverberation of an assembly whose intrinsic positive motive force is sufficient to cross its own motive threshold; in this case, approach behavior will be elicited according to the law of the back-flow of motive force (VIII) and according to the law of approach performance (IXa, ii). In the aversion case, it consists in the firing of an assembly whose negative motive force is sufficient to cross its own motive threshold. In this case, avoidance behavior will be determined jointly by the firing negative assembly and a less negative (or positive) reverberating successor. This determines behavior in the direction of the less negative successor according to the law of avoidance performance (IXa, i).

Sign-gestalt: this term is defined as the knowledge that a sign followed by a direction distance will lead to a significate, e.g., the knowledge that in the presence of A , R_1 leads to B . Our mechanical analogy for the sign-gestalt is two cell assemblies joined through a response control unit by an associational relation. The sign is the antecedent; the direction distance is the response control unit; the significate is the successor.

Sign-gestalt-expectation: this term refers to the expectation that a certain direction distance will lead to the significate; the expectation results from the fact that the sign is presented and perceived. Our mechanical analogy derives from the postulate that the firing of the antecedent arouses reverberation of the successor by the law of conditioned stimuli (VII). That is, if an associational relation joins a and b through r_1 , then a 's firing arouses reverberation (expectation) of b .

Sign-gestalt-readiness: this term refers to a want for some means object by virtue of its instrumental relation to a demanded object. Our mechanical analogy here is the reverberation of a cell assembly whose intrinsic motive force is not sufficient to cross its own motive threshold. It requires reverberation of a successor (VIII) and presentation of the conditioned stimulus of the assembly in question (Vb). The reverberating successor will add a component of motive force to the assembly in question; thus the combined motive force of the assembly will be above threshold, and the conditioned stimulus will arouse reverberation. At this point the assembly in question will function as though it were a "demand." However, termination of its reverberating successor will terminate its own demand characteristics, as its instrumental motive force supply will be cut off.

Sign-gestalt learning: Tolman's theory of learning is briefly the following. In any given training sequence, the subject learns new sign-gestalts, depending on what he perceives. For example, first the animal is in the presence of stimulus A . On Tolman's theorem of the selectivity of perception, the subject will perceive A provided that it is relevant to some present demand (6, pp. 35 and 386). Second, the subject adopts a direction distance R_2 ; that is, he performs behavior R_2 . Third, when the behavior is done, he is in the presence of stimulus B . He will perceive B provided that it too is relevant to some one of his present motives. If the subject has perceived both the antecedent A and the outcome B , then a new sign-gestalt is learned in the performance process; it is that in the presence of A , R_1 leads to B .

Implicit in this description of sign-gestalt learning there is a premise that comes into superficial conflict with Tolman's (6, pp. 343-344) attack on the

law of effect. The point is this: if the outcome B must be perceived in order for learning to occur, and if perception is contingent on motivational relevance, it follows that the outcome B must be either a goal or an instrumentality, a reinforcer or a secondary reinforcer, in order for learning to occur. But Tolman's attack on the law of effect suggests that possibly there is no need of B being a reward for learning to occur (6, p. 343). In justice we must say that Tolman (6, pp. 386-387) recognizes this superficial conflict, but he does not explicitly resolve the confusion. Our mechanical model does, and thus it provides a basis for reorienting the so-called "latent-learning" controversy (see Thistlethwaite, 5) as we will show in a moment.

LEARNING REQUIRES REINFORCEMENT

Our mechanical analogy for sign-ge-stalt learning derives from the learning law of association (VI). Two assemblies become related by an associational relation if a fires, then r_1 is activated, then b fires. But the conditions for the firing of a and b are outlined in the law of assembly activation (V). Both the motive threshold and the stimulus threshold of firing must be crossed before firing will occur. But in order for the motive threshold to be crossed, the assembly must have either sufficient intrinsic motive force (in which case its stimulus is a reinforcer) or sufficient instrumental motive force (in which case its stimulus is a secondary reinforcer). Thus, there is no learning without reinforcement.

But our model does predict *latent learning* provided the B stimulus is a reinforcing stimulus. For, a change in the combined motive force of b can be immediately reflected in two other changes: (a) a change in the combined motive force of a and (b) a change in

the likelihood of the activation of r_1 while a is firing, both without any repetition of the $A-R_1-B$ sequence. This derives from the law of the backflow of motive force (VIII) and from the law of performance (IX). The implication is that a change in the value of the outcome B will change the value of the antecedent A and the likelihood of the response R_1 to stimulus A without any repetition of the $A-R_1-B$ sequence. Thus, learning which was *latent* when the combined motive force of b was insufficient to evoke performance will become evidenced when the combined motive force of b is changed by some operation.

Our suggestion vis-à-vis the rather large experimental program which has centered around the latent-learning controversy is this: experiments which succeed in making the outcome B sufficiently neutral with respect to the present motivational state of the subject will not give evidence of latent learning. We may just as well stop looking for learning without any positive or negative reinforcement, for in these cases the outcome will not be "perceived."

Experiments will demonstrate latent learning, however, whenever the outcome is made motivationally relevant in a positive or negative direction during learning, if the motivational relevance is reversed (as from positive to negative) after training without any further repetitions of the training sequence. In these cases, there will appear (if enough subjects are run) first-trial evidence of changes in response likelihood; such first-trial changes cannot be predicted by Hull's theory. Tolman and Gleitman (7) have reported such an experiment and it has sustained this prediction.

In summary, further experiments should show two things: (a) after $A-R_1-B$ training with a reinforcing stimulus B , changes in the value of B will

be reflected immediately in changes in the likelihood of the $A-R_1$ sequence without any further $A-R_1-B$ sequences required to mediate this change in likelihood; but (*b*) learning will rarely be demonstrated in an $A-R_1-B$ sequence where B has no history as a reinforcer or a secondary reinforcer, or where B is completely irrelevant to a strong present motivation, because in these cases B will not be perceived. In the terms of our model, b will not fire.

STIMULUS CONTROL OF IDEAS

The objection has long been made to cognitive theories that they do not genuinely predict behavior because they are unable to specify clearly before the fact the conditions under which the so-called immanent or ideational determinants of behavior will operate.

Our mechanical model for sign-gestalt theory takes a long step toward meeting this objection. The main cognitive determinants in Tolman's system are perceptions, expectations, readinesses, and demands. Tolman groups the first two, but we separate them. Our model specifies stimulus conditions, or operations under the control of the experimenter for the control of each of these cognitive processes.

Let us presume that our subject has been habituated to the sequence $A-R_1-B-R_2-C-R_3-D$. D is a primary goal, and thus this is the paradigm for any regularly repeated stimulus-response sequence eventuating in a goal. The internal organization resulting from the habituation will be $a-r_1-b-r_2-c-r_3-d$. To arouse the "perception of A " we must fulfill the conditions for the firing of a . Stimulus A plus some conditioned stimulus of d will suffice; for A is the unconditioned stimulus of a , and the reverberation of d assures the motivation of a . At the same time, we have fulfilled the conditions for the "expectation

of B ," that is, the reverberation of b . This is because a 's firing provides a conditioned stimulus for b and d 's reverberation provides adequate motivation; therefore b reverberates and B is expected. Although the conditions for the arousal of the "perception of A " are identical with those for the "expectation of B ," the conditions for the termination of these two states are different. Firing of a will cease upon withdrawal of A ; but reverberation of b will tend to continue until the presentation of B produces firing of b . Next, the presentation of a conditioned stimulus for d combined with a conditioned stimulus for c will produce a "readiness for C ." This is because a conditioned stimulus combined with adequate motivation produces reverberation. The readiness will be terminated by presentation of C (which would fire c and thus terminate reverberation) or of D (which would cut off c 's supply of instrumental motive force by terminating the reverberation of d). Finally, it is quite obvious that the presentation of a conditioned stimulus for d arouses a demand for D , and the presentation of D itself terminates that demand.

An experimental program which makes use of some of these specifications will be outlined briefly in the next section.

THE GROWTH OF APPROACH MOTIVES

In conclusion, I am going to suggest briefly an experimental program for the investigation of the growth and decline of secondary approach motives based on the variables derived from the new model.

In the first place, it has been suggested that the intrinsic motive force of an assembly is a joint function of the number of "transmission units" in the assembly and the "motive force" vested in each unit (Xa). The first

problem in growing a motive, therefore, is to get some transmission units into the assembly, i.e., to get an assembly to start with. To do this we must give our subject some commerce with a stimulus, and then assure the firing of the newly formed assembly for some periods of time (Xb). Presume that we want to form a motive directed at stimulus B as a goal. We may form an assembly and assure its firing by habituating our subject to the stimulus-response sequence $A-R_1-B-R_2-C$ in which C is a primary goal. This forms the cell assembly b . We know the conditions for assuring the firing of b , namely, that during the time intervals while B is presented, if c is reverberating, b will be firing. During these periods of firing, b will be recruiting transmission units (Xb) but these units will be losing motive force (Xc). Thus, we are creating a cell assembly but not a motive.

In the future, however, the growth of positive motive force in b will be a joint function of time intervals of reverberation of b (Xd) and the combined positive motive force of b during these intervals (Xe), and the latter will be a function of the positive motive force in c , and the strength of the association between b and c (VIIIb). To accomplish time intervals of reverberation in b we have to stretch out the time interval between $A-R_1$ and the presentation of B ; that is, we have to give the conditioned stimulus which arouses reverberation in b and then delay the unconditioned stimulus which terminates this reverberation. Therefore, we delay the presentation of B with reference to its place in the habituation sequence. This delay should increase the intrinsic motive force in b (Xd), and should result in a measurable increase in the reward value of the stimulus B . Increases in the reward value of B can be measured by changes in the subject's tendency to pursue this stimulus; I will not go into

specific measures at this point, but they have been developed.

To accomplish a high combined motive force in b during intervals of reverberation, we have to assure a strong associational relation between b and c , and we have to make sure that c is reverberating during the delay. To vary combined motive force, then, we can vary the primary goal C , or vary the amount of habituation which establishes the associational relation.

In the future, the decline of positive motive force in b will be a similar joint function of time intervals of firing of b and the intrinsic motive force of b during those intervals of firing. The specific variables here are quite obvious, and I will not detail them here.

Experiments to carry out this program have been designed and some completed. Two experiments investigating motive force in b as a function of the delay of B have shown that after habituation this delay does produce significant motive growth (4). Experiments to test the effects of other variables are in progress.

SUMMARY

A mechanical model for sign-gestalt theory based on Hebb's (1) discussion of the cell assembly has been outlined. The cell assembly is used as the structural model for both "ideas" and "wants"; these two terms are rendered equivalent except that wants tend to have a higher motive force parameter than ideas. Cell assemblies have two kinds of activation, reverberation (corresponding to "thought" or "motivation") and firing (corresponding to "perception" or "gratification").

The model provides for the formation of associational relations among cell assemblies when there is a succession of stimuli in the environment. For example, if the objective stimulus-response sequence is $A-R_1-B$ and so forth, where

A and *B* are stimuli, then an internal associational relation will be formed $a-r_1-b$, where *a* and *b* are cell assemblies, and r_1 a response control unit. After an associational relation has thus been formed between cell assemblies *a* and *b* through the response control unit r_1 , the firing of *a* will tend to arouse reverberation in *b*, and reverberation in *b* (aroused from some other quarter) will add to the motive force of *a* and *a*'s further antecedents. Thus, the associational relation passes stimulation forward from *a* to *b* and motivation backwards from *b* to *a*. Cell assemblies have two thresholds, a stimulus threshold and a motive threshold; both must be crossed simultaneously before any sort of activation will occur. The stimulus threshold may be crossed by either a "conditioned stimulus" (i.e., a firing antecedent) or an "unconditioned stimulus"; with adequate motivation, the former will produce reverberation, the latter will produce firing. The motive threshold must be crossed by the intrinsic motive force of the cell assembly or by a reverberating successor. Action is elicited when the antecedent assembly of a response control unit is firing, and the successor of the same relation is reverberating, and there is a motivational balance across the response favorable to the outcome.

The position adopted here represents an expansion of the position presented by Hebb (1). Hebb conceives facilitation as flowing both ways across an associational relation. However, he does not anywhere explicitly recognize the necessity that one particular kind of facilitation, namely, that which is here called motive force, can be conceived only as flowing from associational successor to associational antecedent if the problem of motivation is to be solved. I do not mean here that time in the central nervous system flows backwards. There is no hocus-pocus

or magic here. My argument is simply that when cell assemblies are established in a communicating chain of circuits by the succession of their stimuli in the environment, then motivational flow will be from the representor of the successor to the representor of the antecedent.

The model is used to provide a reorientation of the latent-learning controversy. Latent learning is predicted in the sense that a change in the value of an outcome will change the likelihood of its preceding responses without further repetitions of the responses to mediate this change of likelihood. But the model fails to predict learning without reinforcement, for a stimulus must have value to be perceived (a cell assembly must have motivation in order to fire). On this basis, a change of focus in latent-learning experiments is suggested.

The model is used further to provide a new basis for research on the question of the functional autonomy of motives. Full-fledged learned drives are predicted, and the variables in their growth and decline are suggested. In general, it is suggested that the firing of an assembly increases the number of transmission units in the assembly, but decreases the motive force allocated to each transmission unit. Thus, it increases the size of the potential motivating unit, but decreases its motive force. Motive force, however, will grow later as a joint direct function of time intervals of reverberation, and instrumental value during those time intervals, and the size of the reverberating cell assembly. Firing will later tend to extinguish the motive force of an assembly.

The implication is that after habituation of a subject to a stimulus-response sequence such as $A-R_1-B-R_2-C$ where *A*, *B*, *C* are stimuli, *C* being a primary reward, then the lengthening of the R_1-B time interval will tend to produce

increments in the intrinsic reward value of the stimulus B , and lengthening of the time interval of presentation of B will tend to produce decrements in this intrinsic value. Experiments validating the first half of this generalization have been performed (4); others are in progress.

REFERENCES

1. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
2. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
3. HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, 57, 173-180.
4. OLDS, J. The influence of practice on the strength of secondary approach drives. *J. exp. Psychol.*, 1953, 46, 232-236.
5. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97-129.
6. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century, 1932.
7. TOLMAN, E. C., & GLEITMAN, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *J. exp. Psychol.*, 1949, 39, 810-819.

(Received April 8, 1953)

THE PLACE OF PHYSIOLOGICAL CONSTRUCTS IN A GENETIC EXPLANATORY SYSTEM¹

GUDMUND SMITH

University of Lund, Sweden

There are various ways of explaining behavior events physiologically. Let us distinguish here between (a) the use of physiological data, or of constructs derived from such data, and (b) the use of constructs which need not necessarily be verified under the microscope or in the EEG. Some of the more advanced psychological theories are based on hypothetical constructs, as, for example, Hebb's theory (4) and Klein and Krech's "conductivity" concept (5, 7). Such brain models seem to serve as substitutes for psychologically defined models partly because their units of analysis are easy to conceptualize, to handle. The present paper is, however, primarily concerned with the first, less sophisticated and more common kind of physiological theorizing in psychology indicating that physiological processes are *the* manifest reality underlying all behavior events. This approach has often been criticized, and we need not repeat the criticism here (8, 11, 14). Instead, the belief that physiological data represent the basis and origin of mental processes will be used here as a convenient starting point for further inquiry into the place and role of physiological constructs in psychology, especially in a genetic frame of reference.

The assumption that physiological facts represent a "basic level" in the individual, the last link in the explanation of mental processes, is part of a more general assumption that behavior data have to be referred directly to

physical objects, inside or outside, in order to be understood at all. As suggested already by Natorp and Cassirer, however, psychology need not adopt this traditional method of physics and physiology but can (and should) adopt a method of its own. The aim of this method is not to make new constructs in the same objectivizing direction as the natural sciences, but to reconstruct physical objects by tracing them back to their origin, the experiencing subject. Instead of using hypostatized constructs, such as body structure and outside objects, as an explanatory basis for mental processes, the psychologist should analyze the constructs themselves with respect to their genesis in mental processes. Consequently, a physical-physiological unit might be regarded as the outcome of a more or less condensed series of behavior events (perception, concept formation, etc.), the early stages of which are the prerequisite for the later, more adapted and objectivized ones.

An explanatory model concerned with physical-physiological categories is a generalized, abstract conception of reality, in many respects the end product of the conceptual development of Western science. Similarly, a physicalistic ("reality-oriented") frame of reference accepted by the individual can be described as the result of a far-reaching emotional-intellectual socialization. Piaget and Rapaport, among many others, follow in detail this development from primary to secondary stages in our cognitive schemata and thought processes, this acceptance by degrees of a common, objective knowl-

¹ The author wishes to express his gratitude to Drs. George S. Klein and Daniel J. Levinson for valuable criticism.

edge, of detours in thinking (9, 10, 12). Let us, therefore, understand physiological constructs or facts as signs of a more or less objectivizing (physicalizing) set or point of view in human beings; let us regard their role as frames of reference for reality-testing in the individual's development, his adaptation to a stabilized world. The proposition is, then, that the physiological "reality" determines behavior, not merely as a number of causal factors behind the "mental surface" but as a conception in the individual himself of human nature, of reality.

Emmert's law, stating that the apparent size of an afterimage varies directly as to the subject's distance from the projection field, may serve as an illustration (13). According to the proposition, it can be assumed that these size relations hold true when the experienced world of an individual (his relevant region) is conceptualized in a physicalistic, "accurate" way. This implies that afterimage and screen must become isolated from each other, the afterimage as a "subjective" and the screen as an "objective" phenomenon. The afterimage, conceived of in this conventional, physicalistic way, is a constant nerve process; the screen is a nerve process changing in inverse proportion to the screen's distance from the eye. Naturally, the subject need not know anything about retinal areas and the like, only the formal differences between the stable (inside) and the changing (outside) reference systems. As pointed out in an earlier discussion on Emmert's law, the arrangements in most afterimage experiments of this type favor an isolation of image and screen, favor the analytic set necessary to diminish size constancy as far as possible (13).

This being true, the relations in the world we perceive must become equivalent with the relations in the physiological schema. When now the screen

is moved to or from the subject the area of stimulated nervous tissue will be extended or diminished—in linear proportion to the distance—but the afterimage (as excited area) remains constant. Hence, the afterimage will be small or large, respectively, as compared with the excited area of the projection screen. And the same relations appear for the perceived world of our "objective" subjects; their afterimages conform to Emmert's law. But as soon as the conceptual schema is less developed, or, as soon as it is different, there will be deviations from the rule. Children, for instance, often think that the afterimage is a real object like the projection screen, i.e., an object the size of which varies in the same way as other external objects. Consequently, their afterimages do not increase or decrease in relation to the screen at various distances but are apparently size-constant (13). In many adults, too, a negative afterimage (or an eidetic image) is first considered to be an object "out there"; not until late in a series of experiments is the size constancy overcome.

Thus, the variations in apparent size of projected afterimages differ among people because the conceptual frames of reference adopted by them are different. While a physiologist would probably prefer to say that the afterimage follows Emmert's law because of an underlying, constant nerve-process (which might be unstable in children and some adults),² the more reasonable explanation, considering the deviations reported above, seems to be that the individual, for some reason or other, has adapted a conceptual schema in full agreement with a world of linear physical relationships. It is now easy to see

² The more advanced physiologizing psychologist would, of course, use a hypothetical variable to explain deviations in the law. This kind of theorizing will be discussed later.

why many a theory bound to manifest physical structures or observed physiological processes succeeds in explaining only specific and limited forms of behavior; and one understands why some physiological psychologists are eager to make "pure perception" the main object of a psychological science. The classical experimental psychology has sometimes been able to explain response only because the (conventional) physicalistic view is generally accepted in our society. Percepts can be considered as representations of behavior events within a more or less normalized framework of external reality, and, therefore, they must partly agree with a popular, physicalistic model of the world.

The assumption that people perceive (behave) according to conceptual patterns as developed in their life history is not new; indeed, it has been stressed by Jackson, Head, Gelb, Stern, Cassirer, and many of their contemporaries, and later by students of perception and personality (6). Studies of cultural factors in perception, as, for example, comparisons of Rorschach responses in Western communities and primitive tribes (2), also tend to support the assumption; religion, customs, prejudices, the whole reality imposed on us by our society seems to determine what we actually perceive.³ Cases of brain-injured people perhaps illustrate the point most clearly. One of Gelb and Goldstein's subjects, for example, did not see a red color as red in general but only as a specific hue related to well-known objects (e.g., strawberry), because his approach was non-symbolic, because his conceptual schemata lacked centers for a categorized perception of color. His vision in a narrow sense was not impaired, however; he was supposed still to have re-

ceptors for red "in general." But the subject himself could not accept this abstraction any more (3). Psychosomatic medicine can furnish us with further data, e.g., the acceptance of a somatic cause of mental troubles may result in somatic symptoms.

Before concluding this discussion let us develop the considerations once again, but now in terms familiar to the traditional psychologist. It is not necessary to avoid the stimulus-response model altogether in order to show why physical facts often fail to explain experience and behavior data. Stimuli (from outside) and physiological processes have been defined above as objective, generalized conceptions of reality as developed in the empirical tradition of natural science. In the stimulus-response model, behavior is influenced by external stimulation as well as internal (body physiology). But the response is not necessarily directly determined by this stimulation and coherent with its properties; it is, instead, an expression of *how* the stimulation has been received and "acknowledged." A behavior event will become the immediate reflection of a physical-physiological process only if this process is conceived of as "reality" by the subject. As soon as the individual's conception of reality is less "objective," less socialized, the response cannot and will not be a mere prolongation of stimulus (inside or outside). The response or behavior, defined as the outward expression of our experienced world or relevant region, has an immediate physical-physiological basis only when this world is a stimulus reality.

The physiological "level" does not represent the origin of a mental development but a stage in it, often (but not necessarily) the end result of the socialization process of thinking (cf. 4). It seems to be meaningless to ask for physiological facts underlying be-

³ The developments in this field have been excellently summarized and commented on by Dennis (1).

havior phenomena of an individual without knowing whether or not he has accepted the generalized cognitive schema to which these facts belong. This might explain why neurological models as described in the introduction had to be extended over the traditional boundaries of a matter-of-fact science in order to cover more than limited forms of behavior. If, for instance, an individual persists in behaving abnormally in spite of the fact that all known neurological functions in him seem to be normal, if he refuses to adopt the neurologist's reality and is solely governed by his own "unsocialized" experience, it becomes necessary to introduce a hypothetical construct (e.g., integration of brain processes), the derivations of which should be able to explain all behavior deviations, even those without a known physical basis or with an imagined one. This means that neurological constructs in psychology must be more concerned with the reality represented by the wide developmental range of human experience (behavior) than with the limited reality of manifest physical facts and physiological observations, i.e., these hypothetical constructs must remain basically psychological in spite of the physiological language.

The empirical question is, however, how the generalized behavior has developed in different individuals, or why it has developed in some individuals but not in others. A physicalistic schema as accepted by the individual thus gets a personal significance; it may, for instance, be looked upon as a communication or defense mechanism, as a cognitive style, etc. (17). The physiological conception of the world, the impersonal behavior, like all behavior phenomena, ought to be genetically explained (16).

REFERENCES

1. DENNIS, W. Cultural and developmental factors in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 148-169.
2. DU BOIS, CORA. *The people of Alor*. Minneapolis: Univer. of Minnesota Press, 1944.
3. GELB, A., & GOLDSTEIN, K. Psychologische Analysen hirnpatologischer Fälle. X: Ueber Farbennamenamnesie. *Psychol. Forsch.*, 1924, 6, 127-186.
4. HEBB, D. *The organization of behavior*. New York: Wiley, 1949.
5. KESSEN, W., & KIMBLE, G. A. "Dynamic systems" and theory construction. *Psychol. Rev.*, 1952, 59, 263-267.
6. KLEIN, G. S., & KRECH, D. The problem of personality and its theory. *J. Pers.*, 1951, 20, 2-23.
7. KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, 21, 118-148.
8. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.
9. PIAGET, J. Principal factors determining intellectual evolution from childhood to adult life. In *Factors determining human behavior*, Harvard Tercentenary Publ. Cambridge: Harvard Univer. Press, 1937. Pp. 32-48.
10. PIAGET, J. *La naissance de l'intelligence chez l'enfant*. Neuchâtel: Delachaux & Niestle, 1948.
11. PRATT, C. C. *The logic of modern psychology*. New York: Macmillan, 1939.
12. RAPAPORT, D. Toward a theory of thinking. In D. Rapaport (Ed.), *Organization and pathology of thought. Selected sources*. New York: Columbia Univer. Press, 1951. Pp. 689-730.
13. SMITH, G. *Psychological studies in twin differences. With reference to after-image and eidetic phenomena as well as more general personality characteristics*. Lund, Sweden: Gleerup, 1949.
14. SMITH, G. *Interpretations of behavior sequences. With respect to a radical change in the objective situation*. Lund, Sweden: Gleerup, 1952.
15. SMITH, G. Sprache und Erlebnis. *Theoria*, 1952, 18, No. 1, 78-86.
16. SMITH, G. Development as a psychological reference system. *Psychol. Rev.*, 1952, 59, 363-369.
17. SMITH, G., & KLEIN, G. S. Cognitive controls in serial behavior patterns. *J. Pers.*, in press.

(Received April 8, 1953)

A NOTE ON STIMULUS INTENSITY DYNAMISM (*V*)

FRANK A. LOGAN

*Institute of Human Relations, Yale University*¹

In the recent version of his theory (4), Hull postulates as an intervening variable, stimulus intensity dynamism (*V*), which is defined as a function of the intensity of the stimulus and which enters multiplicatively into the determination of excitatory potential. The choice of a theoretical assumption is, of course, the right of the theorist so long as useful predictions follow. However, this paper will attempt to show how Hull might have deduced the relevant empirical phenomena from his theory without the use of *V*.

There are four general data areas for which *V* was especially designed. Let us summarize these and then propose an alternative description.

The first area is the classical conditioning situation where, for example, an increase in illumination is followed by a UCS. If two groups of subjects are exposed to this situation, where all known relevant variables are identical with the exception of the intensity of the CS (i.e., the amount of increase in brightness) the probability of the CR is greater for the group with the more intense CS (e.g., 5).² Hull would deduce this result on the basis of the difference in *V* between the two groups.

Let us, however, recognize that, between trials, the subject is in the con-

textual environment (S_{ee}) containing a dimly illuminated disk, and that any occurrence of the response to S_{ee} is not reinforced by the UCS. When the organism is in the more brightly illuminated environment (S_1), the occurrence of the response is repeatedly rewarded. The situation becomes a discrimination problem in which reinforcement follows the response to S_1 but not to S_{ee} . For a second group of subjects, also nonreinforced for responding to S_{ee} , the rewarded stimulus complex is a still more brightly illuminated environment (S_2). It follows that, since the difference between S_{ee} and S_2 will be greater than the difference between S_{ee} and S_1 (assuming that similarity is a monotonic function of stimulus intensity), there will be greater generalization of the inhibition conditioned at S_{ee} to S_1 than to S_2 . Hull has provided the derivation that the net discriminatory excitatory tendency ($s\dot{E}_R$) will be greater at the positive stimulus the greater the difference between the two stimuli. Therefore $s\dot{E}_R$ will be greater for the group with S_2 as the CS than for the group with S_1 ; a greater probability of CR is expected at the stronger stimulus.

In this derivation, we assumed that the CS represented an increase in the intensity from a zero or minimal value. If, however, the CS were a decrease (so that, between trials, the illumination would be brighter than any stimulus value used), the analysis here presented would lead to the expectation that a group with the more intense CS (but a smaller change from the intertrial situation) would perform more poorly than a second group

¹ The writer is indebted to Drs. Mark A. May, Neal E. Miller, and Burton S. Rosner for a preliminary reading of the manuscript. A research project designed to test quantitatively several of the differential predictions herein contained is being supported by a grant from the National Science Foundation.

² This generalization is not unequivocally supported (e.g., 2). Kessen (5) has suggested a possible analysis of the conflicting results.

with a weaker CS. That is, the non-reinforced S_{ce} in the derivation would be at the upper end of the intensity continuum, and inhibition would generalize down toward the less intense values. It would also be possible to have the stimulus at some intermediate value between trials, and for one group to have a lower intensity serve as the CS, and for another group, a higher intensity. Stimulus intensities appropriately chosen as equal j.n.d. distances away from the intermediate stimulus should give the same probability of CR even though one is more intense than the other. The postulates containing V would be forced to deduce that the difference between the groups favoring the more intense CS would still obtain even under these diverse conditions.

The second set of data for which Hull has found it expedient to employ V concerns those experiments dealing with the time interval between the CS and the UCS. These data suggest that optimal conditioning will obtain at some asynchronism around one-half second, and that intervals either longer or shorter are less effective (e.g., 6). For this reason, Hull's system postulates a stimulus trace which changes as a function of time, and for which a molar stimulus equivalent is calculated for substitution in the equation for V . The trace represents a changing dynamism, and the level of conditioning is assumed to depend upon the V occasioned by a trace of the appropriate age.

The present position would suggest a somewhat different interpretation of the stimulus trace: the numerical value of the trace describes the degree to which a trace of that age represents a change from the conditions of stimulation prior to the onset of the stimulus. The trace function thus states that the onset of a stimulus

produces a continuous change in the stimulus complex, rising rapidly to a maximal difference at about one-half second, and thereafter being reduced until the stimulus complex is, effectively, as it was prior to stimulation. The discrimination learning paradigm again argues that the degree of conditioning will be directly related to the difference between S_{ce} and the CS, where this difference is partially a function of the time since the onset of the CS.³

The third general class of empirical phenomena for which V is directly applicable refers to primary stimulus generalization along intensity as a continuum. Let S_{ce} , S_1 , and S_2 be as above, and choose another stimulus intensity (S_3) which is (a) more intense than S_2 , and (b) equal j.n.d. steps away from S_2 as is S_1 . After the CR is established to S_2 , generalized response tendency is obtained at S_1 and S_3 . Assuming that equal j.n.d. separation means equal difference, habit generalization from S_2 should be the same to each of the two test stimuli. However, it is found that the response strength is greater at S_3 than at S_1 (e.g., 3) which Hull would deduce on the basis of the difference in V .

If the same conceptualization as presented above is followed, the original learning involves a discrimination between S_{ce} and S_2 ; the inhibition at the former will generalize not only to S_2 but to other stimulus intensities. Since we have assumed that similarity is a monotonic function of intensity, S_1 will be more similar to S_{ce} than will S_3 ; S_1 will therefore receive greater

³ Hull's postulate of the stimulus trace (s) does not contain the intensity (S) of the stimulus, but only time (t) since its onset. The most useful interpretation is that s is a calculational device so that $V = f(S, t)$. The analysis offered in this paper would favor the postulate $s = f(S, t, S_{ce})$.

generalized inhibition from S_{ee} than will S_3 . Thus, although S_1 and S_3 will each receive equal generalized habit from S_2 , $s\dot{E}_R$ will be greater at S_3 than at S_1 because there will be less generalized inhibition opposing it; a greater probability of CR is therefore expected at the stronger generalized stimulus.

According to the derivation given here, if the CS were a decrease in intensity (a strong intertrial stimulus), then greater generalized response strength should occur to a stimulus of weaker intensity than to a stimulus equally different from the original CS but stronger. Here, the implication of V is diametrically opposite.

The fourth and final general class of phenomena which involves the use of V occurs in a simple discrimination between S_1 and S_2 (used as above) obtained by the single presentation method. An organism is placed in a starting box and, shortly thereafter, a guillotine door is raised exposing a hinged door of either light or dark gray. This changes the stimulus complex into either S_1 or S_2 , in only one of which locomotion through the door is rewarded. The response strength to the positive stimulus is found to be greater following the discrimination training if the more intense of the pair has been the positive stimulus (e.g., 1). Hull has derived this result on the basis of the greater V at the more intense stimulus.

If, however, S_{ee} is also considered, it will be immediately seen that, when S_1 is the reinforced stimulus complex, both S_{ee} and S_2 will be accruing inhibition which will generalize upon S_1 from both sides. When, however, S_2 is the positive stimulus, it will receive the same amount of inhibition generalized from S_1 as, in the reverse case, S_1 received from S_2 ; but the generalized inhibition from S_{ee} will be less to

S_2 as the positive stimulus than was the case to S_1 as the positive stimulus. Since S_2 would therefore receive the less total generalized inhibition were it the positive stimulus, $s\dot{E}_R$ would be greater when S_2 is positive than when S_1 is positive.

It should be possible to employ single presentation discrimination learning, but to insure that the subject never experiences the contextual environment of the stimulus except at times when either the positive or negative stimulus is present. This would preclude the development of inhibition of S_{ee} . Under such conditions, the analysis followed here would deduce that $s\dot{E}_R$ would be identical whether the weaker or the more intense of the pair was the positive stimulus.

The derivations followed above have been more substantive than exact on the assumption that anyone familiar with the theory will have sufficient facility with its application to discrimination learning to follow the sketch presented. It will be immediately apparent that this discrimination analysis leads to similar deductions as obtained by the use of V if three assumptions are fulfilled: (a) the subject is exposed to the contextual environment of the relevant stimulus, that (b) during such exposure there is a zero or minimal intensity of that relevant stimulus, and that (c) any performance of the response during these intertrial conditions is nonreinforced. Differential implications have been suggested if these assumptions are not met.

While the writer is not aware of research bearing directly upon these implications, several incidental findings seem to favor the present analysis. A number of experimenters have used the offset of a tone as a CS, obtaining satisfactory conditioning even

though dynamism would be near zero. Since V is assumed to enter multiplicatively in determining excitatory potential, it would force sE_R to zero and predict no conditioning. Also it is common, though typically unreported, experience to observe the occurrence of the response between trials more frequently early in training than later. This would be consistent with the hypothesis that the response becomes extinguished to the contextual intertrial stimulus conditions.

Subsequent experimentation⁴ may suggest, of course, that both the above analyses are necessary; that is, that there is an effect determined by the absolute intensity of the CS over and above the effect of the difference between the CS and the intertrial

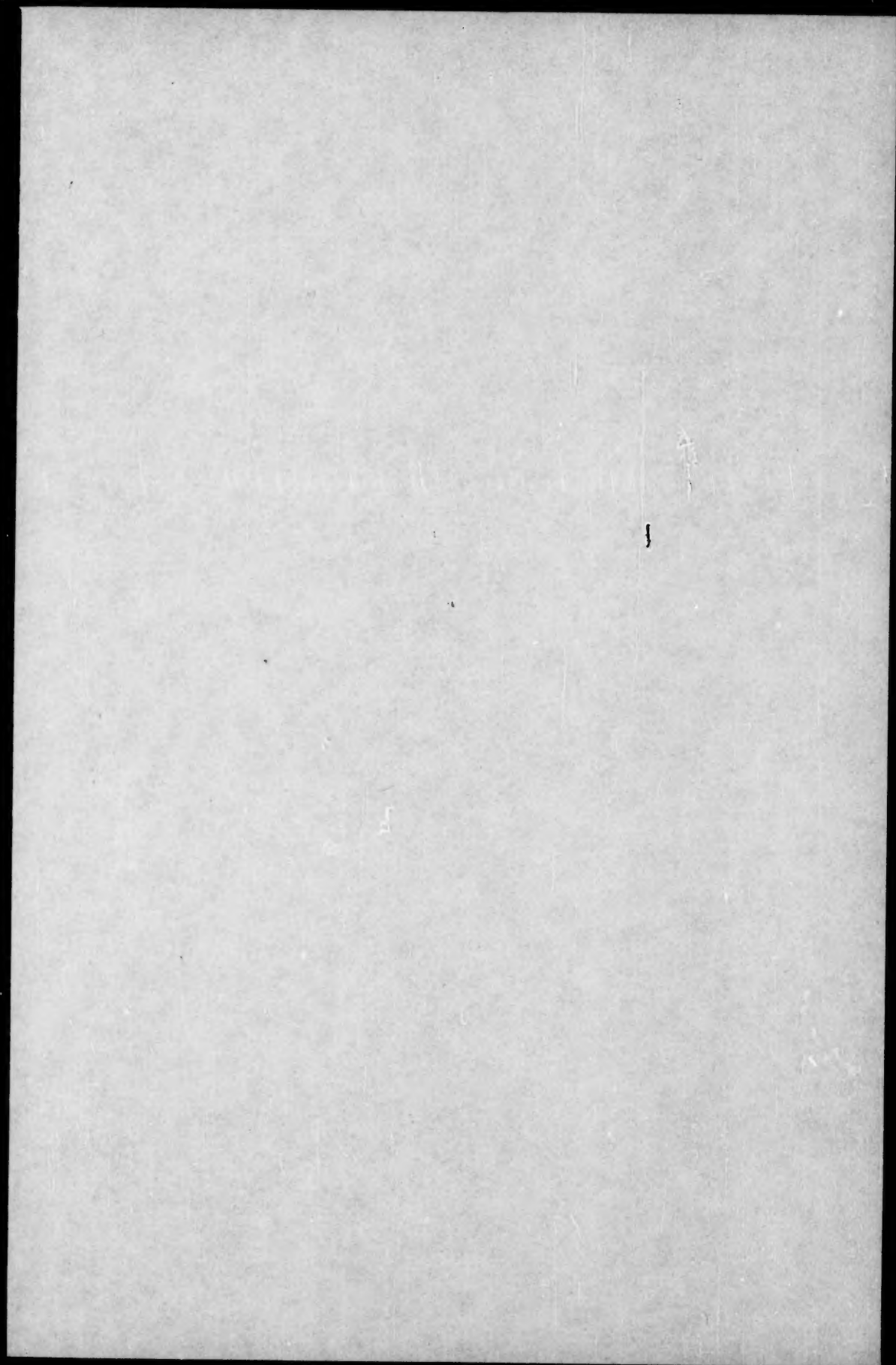
stimulus. More exact research is required before an adequate formulation can be stated.

REFERENCES

1. ANTOINETTI, J. A. The effect of discrimination training upon generalization. Unpublished manuscript, 1950. (Quoted in Hull, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.)
2. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35-40.
3. HOVLAND, C. I. The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279-291.
4. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
5. KESSEN, W. Response strength as a function of conditioned stimulus intensity. Unpublished doctor's dissertation, Yale Univer., 1952.
6. REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.* 1945, **35**, 15-30.

(Received for early publication September 10, 1953)

⁴Subsequent to the preparation of this manuscript, Marvin Schwartz has obtained unpublished data suggesting that a weaker CS is more effective than a stronger one when the contextual intertrial stimulus is intense, and that the occurrence of the response between trials becomes less frequent with practice. The writer has also learned by personal communication that Dr. Charles C. Perkins, Jr. has independently obtained comparable results.



THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by D. W. HARDING

Vol. XLIV. Part 3 August 1953 12s. 6d. net.

OBITUARY NOTICE. David Katz.

C. A. MACE. Homeostasis, needs and values.

W. M. O'NEIL. Hypothetical terms and relations in psychological theorizing.

W. KENNETH RICHMOND. Educational measurement: its scope and limitations. A critique.

L. W. SHEARS. The dynamics of leadership in adolescent school groups.

F. H. GEORGE. 'Either-or' questions in series.

F. A. CHRENKO. Probit analysis of subjective reactions to thermal stimuli—a study of radiant panel heating in buildings.

G. ROBERT GRICE. Hunter's test of the absolute and relative theories of transposition.

IAN M. L. HUNTER. Reply to Professor Grice.

PUBLICATIONS RECENTLY RECEIVED.

Vol. XLIV. Part 4 November 1953 12s. 6d. net.

K. R. L. HALL. Studies of cutaneous pain: a survey of research since 1940.

D. E. BROADBENT. Noise, paced performance and vigilance tasks.

J. A. DEUTSCH. A new type of behaviour theory.

A. C. MUNDY-CASTLE. Electrical responses of the brain in relation to behaviour.

MUKHTAR HAMZA. The dynamic forces in the personalities of juvenile delinquents in the Egyptian environment.

F. V. SMITH, W. SLUCKIN and D. GRAHAM. The efficiency of differently constituted groups of children in different types of tasks.

A. H. D. TOZER and H. J. C. LARWOOD. An analysis of intelligence test scores of students in a university department of education.

KATHLEEN P. WATTS. Influences affecting the results of a test of high-grade intelligence.

CYRIL A. ROGERS. The structure of verbal fluency.

GEORGE HUMPHREY. Five years in the Oxford Chair.

PUBLICATIONS RECENTLY RECEIVED.

The subscription price per volume, payable in advance,
is 40s. net (post free). (U. S. \$6.50).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

Bentley House, Euston Road, London, N. W. 1